HIGHWAY RESEARCH RECORD

Number 409

Traffic Flow Characteristics and Models

7 reports
prepared for the
51st Annual Meeting

Subject Areas

53 Traffic Control and Operations
54 Traffic Flow

HIGHWAY RESEARCH BOARD
DIVISION OF ENGINEERING NATIONAL RESEARCH COUNCIL
NATIONAL ACADEMY OF SCIENCES—NATIONAL ACADEMY OF ENGINEERING

Washington, D.C. 1972

NOTICE

The studies reported herein were not undertaken under the aegis of the National Academy of Sciences or the National Research Council. The papers report research work of the authors done at the institution named by the authors. The papers were offered to the Highway Research Board of the National Research Council for publication and are published herein in the interest of the dissemination of information from research, one of the major functions of the HRB.

Before publication, each paper was reviewed by members of the HRB committee named as its sponsor and was accepted as objective, useful, and suitable for publication by NRC. The members of the committee were selected for their individual scholarly competence and judgment, with due consideration for the balance and breadth of disciplines. Responsibility for the publication of these reports rests with the sponsoring committee; however, the opinions and conclusions expressed in the reports are those of the individual authors and not necessarily those of the sponsoring committee, the HRB, or the NRC.

Although these reports are not submitted for approval to the Academy membership or to the Council of the Academy, each report is reviewed and processed according to procedures established and monitored by the Academy's Report Review Committee.

ISBN 0-309-02081-6
Library of Congress Catalog Card No. 72-9488
Price: $2.60
Available from
Highway Research Board
National Academy of Sciences
2101 Constitution Avenue, N.W.
Washington, D.C. 20418

CONTENTS

FOREWORD

A knowledge of the characteristics of traffic flow and the ability to simulate traffic flow as a tool in developing better control procedures continue to attract relatively wide research support. The 7 papers in this RECORD advance the understanding of stream flow, corridor flow, and network circulation. Researchers, flow theorists, traffic control strategists, and traffic operations authorities should find information in this RECORD helpful.

Munjal, Hsu, and Carpenter report in the first paper about their incorporation of a modified desired-speed probability density into the Boltzmann type of statistical model for multilane traffic. They found that validation results were improved and the model fitted experimental speed data within the margin of error of the data.

Using data taken from aerial photography of freeway traffic, Pahl studied lane-change frequencies on 8-, 6-, and 4-lane freeway sites near off-ramps under various levels of flow. As expected, the greatest frequency of lane changes is to the right for exiting vehicles. The data also showed increased frequency of lane changes to the left by through vehicles that moved back to the right near the off-ramp where exiting vehicles leave the freeway. A discussion of the theoretical and practical implications of Pahl's work by Worrall is also included.

The authors of the next paper analyzed 2 traffic-circulation patterns and traffic-signal timing alternatives using SIGOP. Flynn and Siu conclude that it is practical, with certain minor program modifications, to use SIGOP to improve traffic conditions in a grid street network.

Lieberman describes development of the SCOT model, an evaluative and design tool to predict the performance of alternative control policies and freeway configurations prior to field implementation. The model allows for evaluation of flow in an entire corridor network system, inlcuding the freeway, ramps, frontage roads, and parallel and feeder arterials.

The UTCS-1 network simulation model is described by Lieberman, Worrall, and Bruggeman. The model is called an evaluative tool for urban traffic control policies, and the paper includes descriptions of some simple demonstrations.

Again, taking data obtained by aerial photography, Pahl studied gap-acceptance maneuvers by exiting and through vehicles changing lanes on 8-, 6-, and 4-lane freeways under various flow conditions. His review of the data also included attempts to quantify accident risks assumed for themselves, and imposed on others, as drivers accepted gaps in changing lanes. Risks are reported to vary between exiting and through lane-changers, and exiting drivers are shown to find smaller gap sizes acceptable as they approach their desired off-ramp.

In the final paper, Munjal and Hsu discuss what they describe as the merits and shortcomings of major traffic control concepts and algorithms, from isolated intersections to arterials and networks. It is considered to be a rather concise summary of the state of the art in traffic signal network control.

EXPERIMENTAL VALIDATION OF MODIFIED BOLTZMANN TYPE OF MODEL AND SHIFT MODEL FOR MULTILANE TRAFFIC FLOW

P. K. Munjal, Y. S. Hsu, and R. Carpenter, System Development Corporation, Santa Monica, California

The Boltzmann type of statistical models proposed by Prigogine et al. for time-independent, space-homogeneous, multilane traffic assumes that traffic flow is described by 3 processes—interaction, relaxation, and adjustment—and has been shown earlier to produce poor validation results. In this paper we propose a new model that uses a shifting process $f^{*}(v) = \gamma f^{0}(\gamma v)$, where γ is a concentration-dependent parameter that replaces the desired speed-density $f^{0}(v)$ in the relaxation process. The resulting modified Boltzmann type of model is shown to have many desirable properties and fits experimental speed data within the data error margin. A second model, called the shift model, uses only the shifting process and neglects the interaction, relaxation, and adjustment processes of the Boltzmann type of model, gives results only slightly less optimal than the result of the best possible fit obtained by the modified Boltzmann type of model, and is far better than the original Boltzmann type of models. Some applications of the shift model are discussed.

A RECENT paper (1) critically examined the Boltzmann type of statistical models for multilane traffic flow developed by Prigogine et al. (2, 3) and was followed by an experimental validation study (4) of the Prigogine models for time-independent, space-homogeneous, 2-lane, unidirectional traffic flow. It was concluded (4) that the models proposed by Prigogine et al. were incapable of producing a realistic description of the traffic flow behavior and gave poor agreement between the experimental and the computed speed distribution functions even under low-to-moderate (i.e., 49.5 and 88.4 vehicles/mile for a 2-lane highway) traffic conditions. This validation study also indicated that the addition of an adjustment term of the form suggested by Prigogine et al. led to no significant improvement in the realism of the model.

From the validation study it seems that Prigogine et al. were unable to incorporate dominant factors of traffic dynamics into their models. When traffic concentration increases, the Prigogine models predict a "pileup" (through the interaction process) behind the vehicles moving at the minimum desired speed (4), while it is evident from observations that many vehicles are now traveling below minimum desired speed.

In the basic Boltzmann type of model (2), the change of speed was assumed to be due to 2 processes: the interaction process and the relaxation process. Although the interaction process was derived analytically, it was argued (1) that it overestimated the interaction effect.

The generalized Boltzmann type of model (3) with an additional adjustment term seems of little importance from the earlier validation study (4). The adjustment term relaxes some vehicles to the average speed in the relaxation process rather than to their desired speeds. That is, the speed-density shows a δ function (physically, we mean that the number of vehicles traveling in a specific speed range, say 3 to 5 ft/sec, is substantially higher than those immediately outside of that range) at $v = \bar{v}$, the average speed. However, the fact that the validation study (4) shows no significant improvement can be interpreted to mean that the assumption of being relaxed to the average speed is physically ungrounded. Furthermore, in no way does it solve the pileup problem.

The Boltzmann type of models shows even worse results in high concentration, in that the speed-density becomes negative for certain ranges of speeds when concentration increases to a certain extent, particularly when the average speed becomes less than the minimum desired speed. Under such circumstances, Prigogine et al. (3) were forced to assume that a fraction of the vehicles had to have 0 speed. The Appendix gives a more detailed description of the basic and generalized Boltzmann type of models.

A clearly unsatisfactory property of the Boltzmann type of models is that the actual minimum and maximum speeds are unchanged from the desired minimum and maximum speeds. This property is reflected in the relaxation process, in that the vehicles are relaxed to their original desired speeds after the interaction with other vehicles.

Our first attempt to modify the Boltzmann type of models is to incorporate a shifting process in the relaxation process. This shifting process shifts the desired speed at higher concentration toward lower speeds, uniformly through a shift parameter γ. We shall call this shifted speed-density the relaxed speed-density. When the desired speed-density in the relaxation process of the basic Boltzmann type of model is replaced by the relaxed speed-density, we have obtained the modified Boltzmann type of model. The parameter γ depends on the overall vehicle concentration rather than on the local concentration. That is, we ignore the spatial correlations among vehicles that exist in real traffic in a manner consistent with that in which the interaction process is handled in a Boltzmann type of model. By doing this, we achieve smaller variation in speeds and lower maximum and minimum speeds when vehicle concentration increases. These properties are desirable.

Our second model uses the shifting process alone and neglects the interaction, relaxation, and adjustment processes in the original Boltzmann type of models. We shall call this simple model the shift model. Validation results of this shift model will give indications of the importance of the shifting process concept, compared to other Boltzmann processes.

As in the previous validation study (4), only the time-independent, space-homogeneous case is considered. Available Federal Highway Administration data are most suitable for this type of validation, and the Boltzmann type of integrodifferential equation can easily be solved for this case.

MODIFIED BOLTZMANN AND SHIFT MODELS

The basic objective of the Boltzmann type of model is to predict the speed-density function f(x, v, t) at any vehicle concentration from a knowledge of the desired speed or free speed density function $f^0(x, v, t)$. It is assumed that this desired speed-density is realized in the limit of very dilute vehicle concentrations.

We define f(x, v, t) = speed-density function of vehicles at time t and point x whose actual speed is v considered with respect to space.

The corresponding desired speed-density function is $f^0(x, v^0, t)$ = speed-density function of vehicles at time t and point x whose drivers have desired speed v^0 considered with respect to space.

In addition, relaxed speed-density is defined as $f^*(x, v^*, t)$ = speed-density function of vehicles at time t and point x whose drivers have relaxed speed v^* considered with respect to space.

We also define v_{max}, v_{min} = upper and lower limits respectively of the actual speeds, and define v^0_{max}, v^0_{min}, v^*_{max}, v^*_{min} similarly.

The space-mean speed can now be defined as

$$\bar{v}(x, t) = \int_0^\infty vf(x, v, t)dv$$

The desired space-mean speed, $\bar{v}^0(x, t)$, and the relaxed space-mean speed, $\bar{v}^*(x, t)$, are defined similarly.

Available Federal Highway Administration traffic analyzer data were measured according to time (4). A transformation between space and time data (5),

$$f_s(v) = (\bar{v}/v)f_t(v)$$

where the subscripts t and s denote time and space respectively, was used to obtain spacewise data. This paper considers only the analysis and validation of speed-density in space; the case of speed-density in time can be similarly carried out.

The relaxed speed-density f^* is hypothesized, being obtained through a shifting process of the desired speed-density.

$$f^*(x, v^*, t) = \gamma f^0(x, v^0, t) \tag{1}$$

and

$$v^0 = \gamma v^* \tag{2}$$

We shall call γ the shift parameter in the sequel. This shift parameter γ is assumed to be a function of concentration c with the following properties

$$\gamma \geq 1, \lim_{c \to 0} \gamma = 1, \text{ and } \lim_{c \to c_j} \gamma = \infty$$

where c_j is the jam concentration.

It is easy to see that $f^*(x, v^*, t)$ is indeed a probability density function as

$$\int_0^\infty f^*(x, v^*, t)dv^* = \int_0^\infty \gamma f^0(x, v^0, t)\ (dv^0/\gamma) = 1$$

since $f^0(x, v^0, t)$ is a probability density function.

As in the previous work (Appendix), the rate of change of f(x, v, t) with time is expressed as

$$[df(x, v, t)]/dt = \{[\partial f(x, v, t)]/\partial t\} + \{\partial f(x, v, t)]/\partial x\} \cdot (dx/dt) \tag{3}$$

and the change of f within time is assumed here through 2 processes: the interaction process and the relaxation process. Here we have neglected the adjustment term because it was shown earlier (1) that the relaxation and adjustment terms could be combined as

$$(\partial f/\partial t)_{rel} + (\partial f/\partial t)_{adj} = -[(f - \tilde{f})/T']$$

where

$$\tilde{f} = \eta f^0 + (1 - \eta)\ \delta(v - \bar{v}),$$
$$\eta = T'/T,$$
$$T' = T/(1 + \lambda T),$$

and T and λ are the relaxation time and the adjustment constant respectively (Appendix). This simply means that the relaxation process relaxes speed back to $\tilde{f}$ rather than f^0. More precisely, η fraction of vehicles relaxes back to f^0, and $(1 - \eta)$ fraction relaxes back to $\bar{v}$. Therefore, we can write Eq. 3 as

$$df/dt = (\partial f/\partial t)_{int} + (\partial f/\partial t)_{rel} \tag{4}$$

$$(\partial f/\partial t)_{int} = (1 - P)(\bar{v} - v)f \tag{5}$$

$$(\partial f/\partial t)_{rel} = - [(f - \tilde{f})/T'] \tag{6}$$

where P = probability of passing. The contribution to the time rate of change of f due to the interaction process is analogously the same as before (1).

The relaxation process in the present model assumes f relaxes back toward f^* rather than toward f^0. That is,

$$(\partial f/\partial t)_{rel} = - [(f - f^*)/T] \tag{7}$$

instead of Eq. 6 in the generalized Boltzmann type of model.

Finally, the interaction term in Eq. 5 and the relaxation term in Eq. 7 are added together to obtain the modified Boltzmann type of equation for traffic flow.

$$df/dt = (1 - P)(\bar{v} - v)f - [(f - f^*)/T] \tag{8}$$

For time-independent, space-homogeneous traffic flow, Eq. 8 reduces to

$$f(v) = [f^*(v)]/[1 - \beta(\bar{v} - v)] = [\gamma f^0(\gamma v)]/[1 - \beta(\bar{v} - v)] \tag{9}$$

where $\beta \equiv (1 - P)T$.

The incorporation of an adjustment term by Prigogine et al. (3) has achieved qualitatively the effect of follow-the-leader type of traffic theory. Because the decrease in speed dispersion (variance) is a major consequence of this adjustment, we can see that it is also achieved through the shifting process, i.e., $var(v^*) = 1/\gamma^2\ var(v^0)$ (for proof, see an earlier paper, 6) in the modified model. Furthermore, the vehicle concentration can be divided into 3 sections, corresponding to cases A, B, and C (Appendix), and the basic and generalized Boltzmann type of models produce realistic speed-densities only for case A. The introduction of γ gives realistic speed-densities for the entire range of vehicle concentration by shifting both maximum and minimum speeds toward lower values, that is, $v_{max} = v^*_{max} = 1/\gamma\ v^0_{max}$ and $v_{min} = v^*_{min} = 1/\gamma\ v^0_{min}$. Therefore, we have achieved the objective of extending the model so that it is applicable over a wider range of vehicle concentrations. The available experimental data support our argument that maximum and minimum speeds decrease at higher concentration. Thus, our proposed modified Boltzmann type of model retains the general Prigogine form but relaxes speed toward f^* rather than f^0. By doing this, we have retained all desirable properties and deleted the shortcomings.

We shall focus our discussions and validations on the modified model (Eq. 9).

Some interesting properties of γ and β are illustrated below.

$$\gamma = \bar{v}^0/\bar{v}^*$$
$$\gamma^n = \overline{(v^0)^n}/\overline{(v^*)^n} \tag{10}$$

for positive integer n.

$$\sigma^2_{v^*} = (1/\gamma^2)\ \sigma^2_{v^0} \tag{11}$$

where $\sigma^2_{v^*}$ and $\sigma^2_{v^0}$ are the variances of v^* and v^0 respectively.

$$\gamma_{max} = \bar{v}^0/\bar{v} \tag{12}$$

and γ takes its maximum value, γ_{max}, when $\beta = 0$.

$$\beta = (\bar{v}^* - \bar{v})/\sigma_v^2 \tag{13}$$

where σ_v^2 is the variance of v.

$$\beta < 1/(\bar{v} - v_{min}^*) \tag{14}$$

The proofs are given in an earlier paper (6).

Some of the important and interesting properties of the Boltzmann type of models have been described in the Appendix. In the following discussion, we shall see that the modified model retains all promising properties and avoids the shortcomings.

The property that $\gamma \geq 1$ implies that $\bar{v}^* \leq \bar{v}^0$, $v_{min}^* \leq v_{min}^0$, and $v_{max}^* \leq v_{max}^0$; and we are now free of the restriction that the actual maximum and minimum speeds do not change from the desired maximum and minimum at very dilute concentration. These differences do exist, as we shall see later in validation. Empirically, we should have $\bar{v} \to 0$ when $c \to c_j$ and $v \to v^0$ when $c \to 0$. This means that in the former case $v_{min} = v_{max} = 0$, an obvious change from the desired minimum and maximum. The modified model gives $\gamma \to \infty$ as $c \to c_j$, or $v_{min}^* = v_{max}^* = 0$, a satisfactory result. In the latter case, the modified model gives $\gamma \to 1$ as $c \to 0$, or $v_{min}^* = v_{min}^0$, $v_{max}^* = v_{max}^0$, again a satisfactory result.

The property $\sigma_{v^*}^2 = (1/\gamma^2)\,\sigma_{v^0}^2$ implies that the variance (dispersion) of v^*, and consequently of v, decreases when traffic concentration increases, a phenomenon observed in the experimental data. This also retains the major merit of the adjustment term in the generalized Boltzmann type of model without the physically unsound condition that a finite percentage of vehicles have to travel exactly at the speed $\bar{v}$.

The role of β, as we examine Eq. 9, is to increase the speed-density at speeds $v < \bar{v}$ and decrease the speed-density at $v > \bar{v}$. This is the same as was observed from the Boltzmann type of models. However, here we have imposed a bound (Eq. 14) on β so that the speed-density will never be negative. This enables us to avoid the unrealistic assumptions that f^0 has a pole at $v = 0$ and is 0 between $v = 0$ and $v = v_{min}^0$.

The relations among the desired speed, the relaxed speed, and the actual speed-density functions are shown in Figure 1. Contrary to the Boltzmann type of models (Fig. 11), this modified model represents the entire $\bar{v}$ - c (speed-concentration) range and does not give poles at $v = \bar{v}$, $v = v_{min}$, or $v = 0$ unless the traffic is at jam concentration, in which case $f(v) = \delta(v)$, or unless the desired speed-density specifies poles at certain v. When concentration increases, v_{min} and v_{max} shown in Figure 1 will shift farther left through a larger value of γ and will not go beyond $v = 0$, in which case $\gamma = \infty$.

A second model, called the shift model, that uses only the shifting process to describe traffic flow and ignores all other Boltzmann type of terms (i.e., interaction, relaxation, and adjustment) is also proposed. This model has the following form:

$$\begin{aligned} f(v) &= \gamma_{max} f^0(\gamma_{max} v) \\ &= (\bar{v}^0/\bar{v})\, f^0[(\bar{v}^0/\bar{v})\, v] \end{aligned} \tag{15}$$

Although the shift model is simple in structure, the desired properties of decreasing minimum and maximum speeds and the speed variance at higher concentrations are achieved. We summarize the properties of the shift model as follows:

1. $\gamma_{max} = \bar{v}^0/\bar{v} \geq \bar{v}^0/\bar{v}^* = \gamma$. $\gamma_{max} \to 1$ when $c \to 0$ $(\bar{v} \to v^0)$ and $\gamma_{max} \to \infty$ when $c \to c_j$ $(\bar{v} \to 0)$.
2. $\sigma_v^2 = (1/\gamma_{max}^2)\,\sigma_{v^0}^2$. The variance of computed speed by the shift model is $1/\gamma_{max}^2$ of the variance of the desired speed. We have, therefore, achieved the decrease in speed variance.
3. $v_{min} = (1/\gamma_{max})\, v_{min}^0$ and $v_{max} = (1/\gamma_{max})\, v_{max}^0$. The minimum and maximum speeds from the shift model are reduced to $1/\gamma_{max}$ of the desired minimum and maximum respectively.

In the next section we shall see that the shift model, which uses only the shifting process (Eq. 15), produces good validation results. On the contrary, the basic and

generalized models that use interaction, relaxation, and adjustment processes without the incorporation of the shifting process produce very poor validation results. This suggests that the shifting process alone plays a much more dominant role in describing traffic flow compared to the interaction, relaxation, and adjustment terms in the Boltzmann type of models.

VALIDATION PROCEDURE AND RESULTS

We attempt to validate the modified Boltzmann type of model and the shift model and compare them with the basic and generalized Boltzmann type of models for time-independent, space-homogeneous conditions and for low-to-moderate vehicle concentrations because (a) the closed-form solution is available in Eq. 9, (b) the available traffic analyzer data represent these conditions, and (c) the basic and generalized Boltzmann equations are promising only for concentration ranges given by case A (Appendix). We are also planning to use aerial photographic data for similar studies so that these models are tested over a variety of sites and vehicle concentrations.

Equations 9 and 15 represent relations between the actual speed-densities at any traffic concentration for the modified Boltzmann and the shift models and the desired speed-density or free speed-density realized at very dilute traffic concentrations. The basic approach is to validate the relation between f^0 and f given by Eqs. 9 and 15. For this purpose, the desired speed-density f^0 is measured, and the other parameters of Eq. 9 are determined; the actual speed-density f can be computed from f^0 for any concentration. To evaluate the validity of the model requires that such computed speed-densities be compared with speed-densities that have been measured at different concentrations.

The validation for Eq. 15 is straightforward, as there is no parameter estimation involved. However, the parameters γ and β in Eq. 9 need to be determined from experimental data. Although β has an analytical expression (Eq. 13)

$$\beta = (\bar{v}^* - \bar{v})/\sigma_v^2$$

it is difficult to obtain its value without knowing $\bar{v}^*$, or equivalently γ, and σ_v^2. Furthermore, the calculation of σ_v^2 requires far more information than the average speed $\bar{v}$. Therefore, we proceed as follows.

Given the model

$$f(v) = \{[f^*(v)]/[1 - \beta(\bar{v} - v)]\} + \epsilon = \{[\gamma f^0(\gamma v)]/[1 - \beta(\bar{v} - v)]\} + \epsilon \qquad (16)$$

Find γ and β such that

$$1 \leq \gamma \leq (\bar{v}^0/\bar{v})$$

$$\int_0^\infty \{[\gamma f^0(\gamma v)]/[1 - \beta(\bar{v} - v)]\}\, dv = 1 \qquad (17)$$

and

$$d^2 = \int_0^\infty \left(f(v) - \{[\gamma f^0(\gamma v)]/[1 - \beta(\bar{v} - v)]\}\right)^2 dv \qquad (18)$$

is minimized. We note that there is one, and only one, value of β besides 0 satisfying the normalization of Eq. 17. Proof is given in an earlier paper (6). The variables

$f(v)$, $\bar{v}$, and $f^0(v)$ are respectively the measured speed-density function, its space-mean speed, and desired speed-density function realized at very dilute concentration. ϵ is introduced here to represent errors caused by measurements, traffic disturbances, and so on. The minimization of Eq. 18 is the standard least squares minimization, and d^2 is a measure of goodness of fit.

The desired speed-density at very dilute concentration f^0, the actual speed-density f, and the space-mean speed corresponding to f are computed from the available traffic analyzer data. There are 5 data sets considered as 5 constant flow levels; these are represented by the 5 circles shown in Figure 2. The flows and concentrations are the sums of 2-lane traffic. The 3 lowest flow levels are used to estimate $f^0(v)$ and the 2 highest flow levels are used for experimental validation.

We first make the following computations:

1. $\hat{f}^0$, estimate of desired speed-density function in space, calculated by $\hat{f}^0 = (f_1 + f_2 + f_3)/3$, where f_1, f_2, and f_3 are the speed-density functions of the lowest 3 flow levels shown in Figure 2. The estimate $\hat{f}^0$ is shown in Figure 3.
2. $\hat{f}$, estimate of speed-density function at a flow above free-flow level. The highest 2 flow levels shown in Figure 2 represent this case. They are 2,345 cars/hour and 3,695 cars/hour corresponding to concentration 49.5 cars/mile and 88.4 cars/mile respectively (for 2-lane). Figure 4 shows an example of $\hat{f}$ for the 2,345 cars/hour flow.
3. $\hat{\bar{v}}$, estimate of space-mean speed corresponding to $\hat{f}$ and computed as the harmonic mean, $1/(1/n \sum 1/v_i)$, of the speeds $v_1, v_2, \ldots, v_n$ of the successive cars passing the observation point.

The value found for $\hat{\bar{v}}^0$, the estimated desired mean speed, was 48.18 mph, and the estimates $\hat{\bar{v}}$ for the 2 flow levels of 2,345 and 3,695 cars/hour were 47.43 and 41.92 mph respectively. Next, various trial values for γ were selected from the range $1 \leq \gamma \leq \bar{v}^0/\bar{v}$, and a corresponding value of β was calculated from Eq. 17.

Each pair of γ and β, along with the estimates given above, $\hat{f}^0$, and so on were then used in the modified Boltzmann type of Eq. 9 to calculate estimated speed-density $\hat{f}$ for each of the 2 flow levels. The number d^2 in Eq. 18 was computed for each choice of parameters γ and β in which we use $\hat{f}$ for f and $\hat{f}$ for $[\gamma f^0(\gamma v)]/[1 - \beta(\bar{v} - v)]$, and the "best" fit was then determined to be given by the values of γ and β that resulted in minimum d^2.

Before discussing these results, let us first examine the criterion on which one can decide whether a particular calculated speed-density is a "good" fit to the experimental density. We first need some measure of the scatter (i.e., fluctuations) in the experimental speed-densities. Because our estimate of the desired speed-density f^0 was obtained by averaging 3 low flow-concentration periods, the sum of the squares of the differences between these 3 low speed-densities (taken pairwise) would be a reasonable measure of the fluctuations of the experimental data. These 3 densities are based on observations of 674, 1,090, and 945 vehicles respectively. The sum of squares, d^2, of density differences was found to be 0.0026 (between 1 and 2), 0.0028 (between 1 and 3), and 0.0035 (between 2 and 3). Next, because the 2 higher flow-concentration periods against which we validated the model are based on observations of 984 and 1,369 vehicles (i.e., numbers of vehicles similar to those in the low-concentration periods), we argue that any model prediction of a speed-density that results in a sum of squares of differences (when compared to the corresponding experimental distribution) of approximately 0.003 or less would be acceptable.

Table 1 gives our findings. First, it is clear that the lower density case is relatively uninteresting. All of the sum of squares of differences are similar and come close to our criterion for significance. Here it makes little difference whether we use the basic Boltzmann model with $\gamma = 1$ and β determined by Eq. 17 or the generalized Boltzmann model (it happened that $\lambda = 0$ gave the best fit) or the modified Boltzmann model with γ and β chosen to give least squares or even the simple shift model that uses γ_{max}. These results for the lower density case are hardly surprising when one refers to data shown in Figure 2. There is very little deviation from a flow versus concentration linear relation (i.e., $\hat{\bar{v}}$ is only 0.75 mph less than $\hat{\bar{v}}^0$). Thus, all of the various models make very minor adjustments on their inputs, $\hat{f}^0$, and return density, $\hat{f}$, that compares some-

Figure 1. Actual, f, desired, f^o, and relaxed, f^*, speed-density functions.

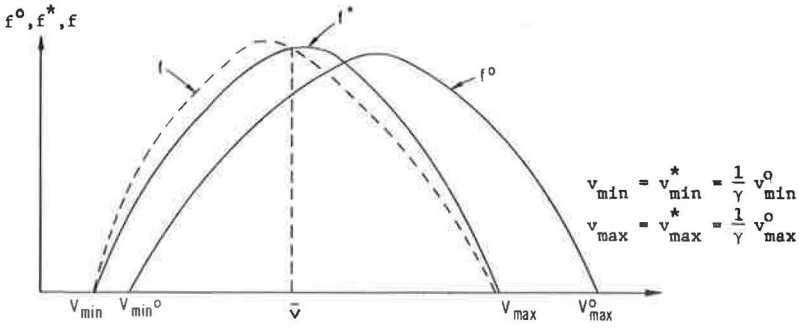


Figure 2. Measured flow and concentration.

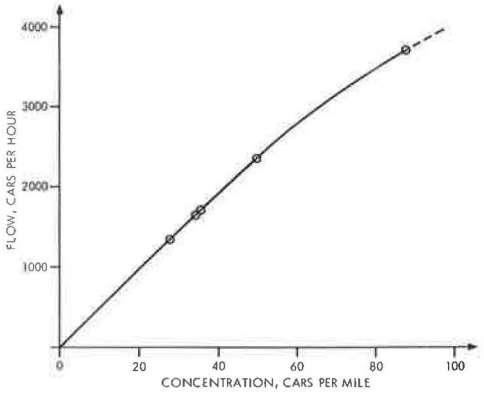


Table 1. Sum of squares of various models.

Model and Number of Vehicles	γ	β	d^{2a}
Basic Boltzmann			
2,345 vehicles/hour	—[b]	0.0175	0.0036
3,695 vehicles/hour	—[b]	0.0861	0.0495
Generalized Boltzmann			
2,345 vehicles/hour	—[b]	0.0175	0.0036
3,695 vehicles/hour	—[b]	0.091[c]	0.0467
Modified basic Boltzmann			
2,345 vehicles/hour	1.010 to 1,016	0.005 to 0	0.0033
3,695 vehicles/hour	1.116	0.036	0.0012
Shift			
49.4 vehicles/mile	1.016	—[d]	0.0033
88.4 vehicles/mile	1.149	—[d]	0.0015

[a] Avg between 3 lowest concentration speed-densities is 0.003.
[b] Does not apply; equivalently, $\gamma = 1$.
[c] $\lambda = 0.622$.
[d] Does not apply; equivalently, $\beta = 0$.

Figure 3. Measured desired speed-density.

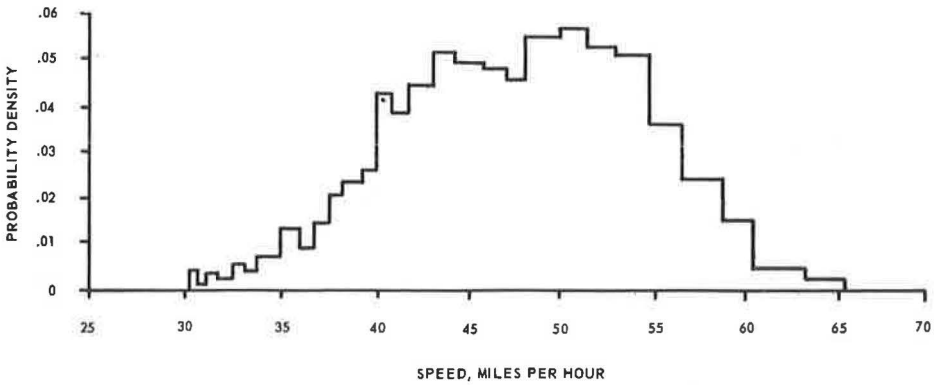


Figure 4. Measured and computed speed-densities by shift model.

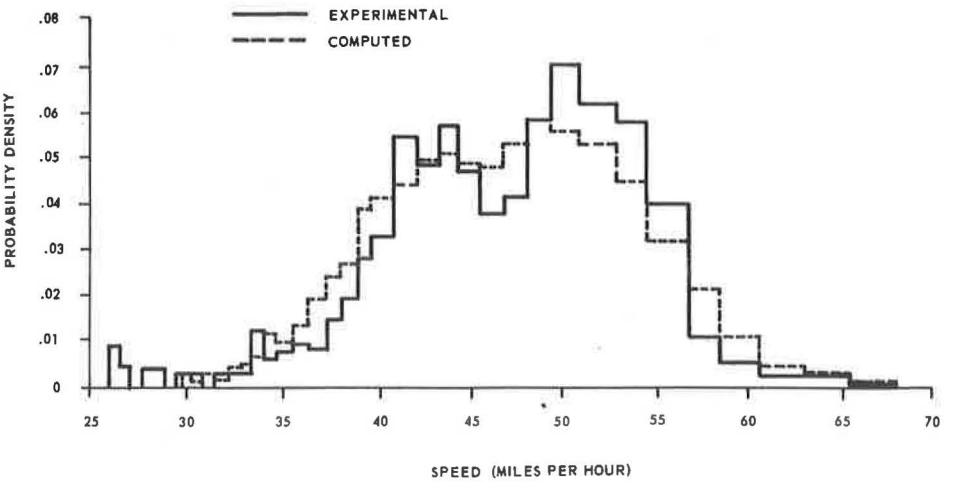


Figure 5. Measured and computed speed-densities by basic Boltzmann.

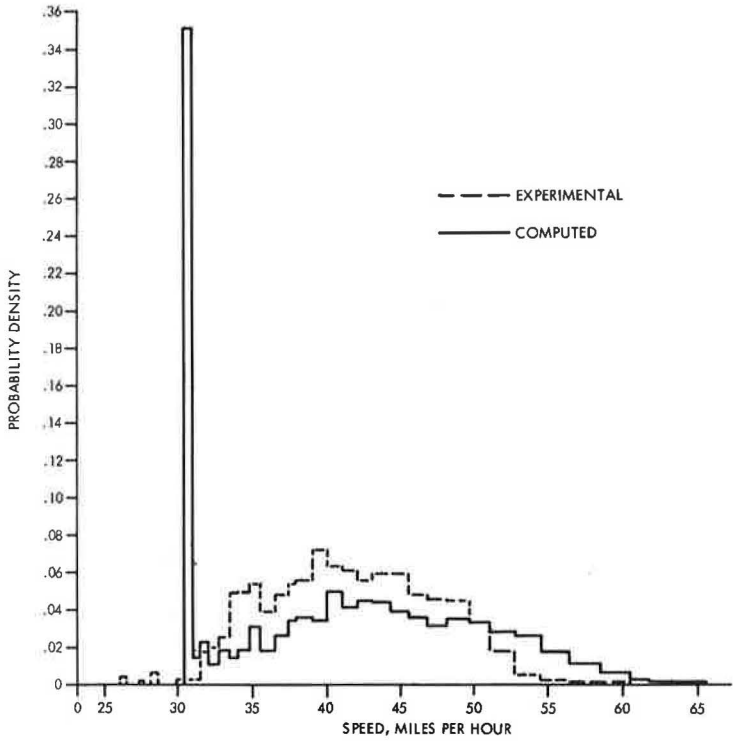


Figure 6. Measured and computed speed-densities by generalized Boltzmann.

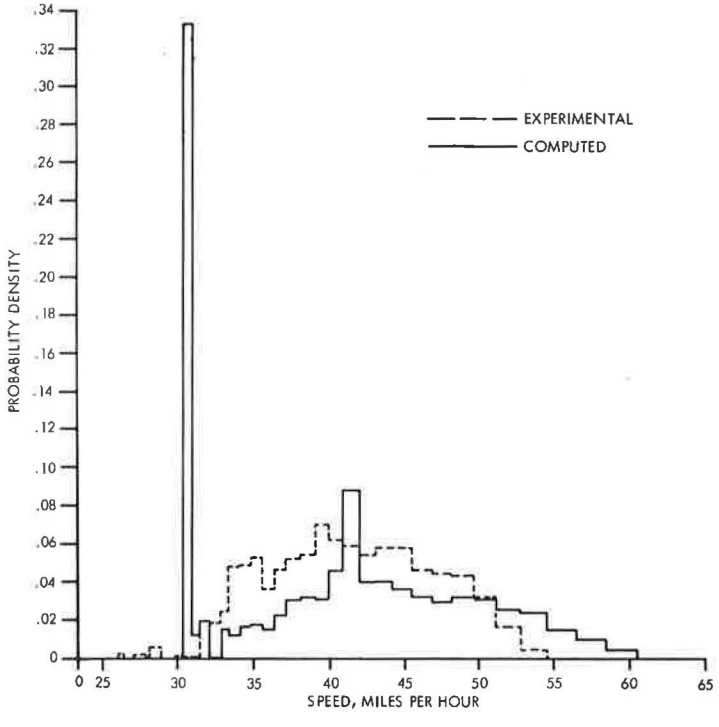


Figure 7. Measured and computed speed-densities by modified Boltzmann.

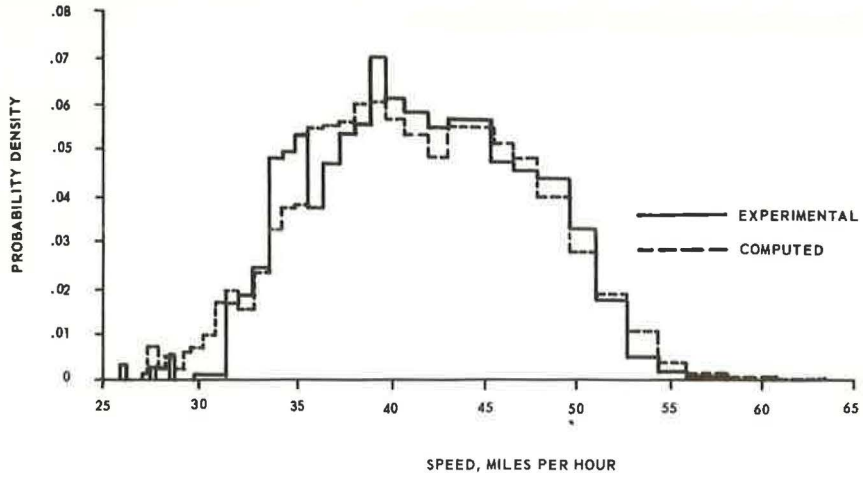


Figure 8. Measured and computed speed-densities by shift model.

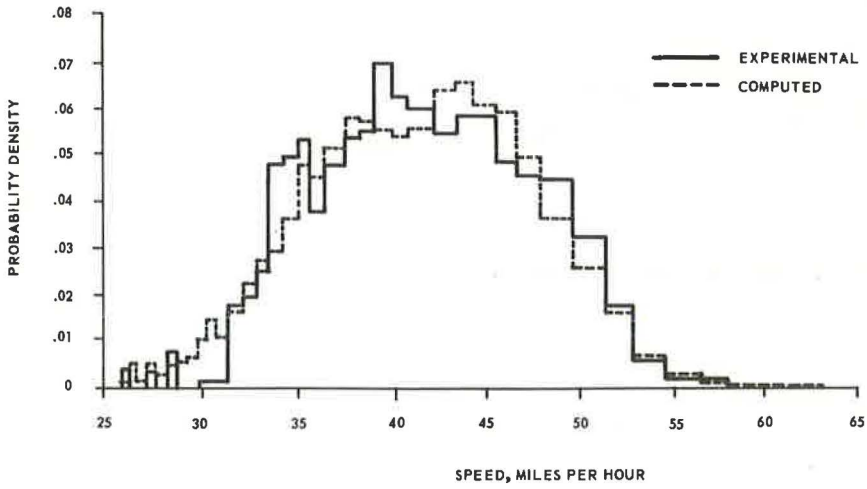
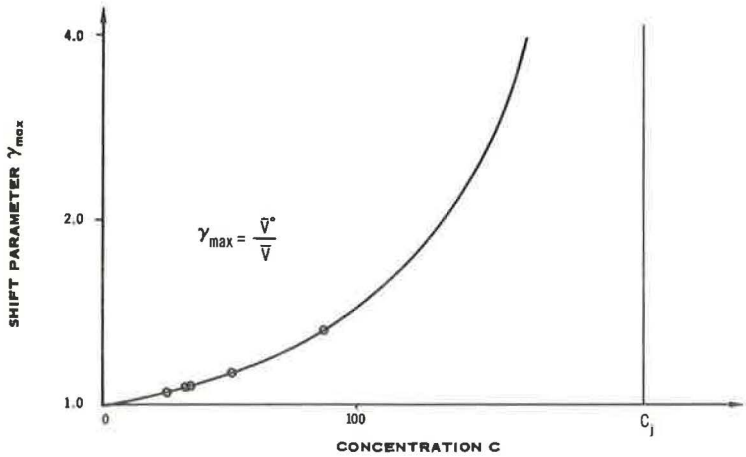


Figure 9. Shift parameter γ_{max} as function of concentration.



what favorably with $\hat{f}$; therefore, in the low-concentration case, these models do not produce significantly different results. We also notice that all of the trial values of γ and β within these ranges gave roughly the same sum of squares, d^2, and that $\gamma = \gamma_{max}$ and $\beta = 0$ yield an acceptable fit even though we are then dealing with only the shifting process. Figure 4 shows the computed speed-density obtained from this shift model. Results obtained by using other models are very similar, and there is no need to show them graphically.

Now let us examine the results given in Table 1 for the higher density and flow case, 88.4 vehicles/mile and 3,695 vehicles/hour (2 lanes). For this case, $\hat{\bar{v}} = 41.92$ mph compared to $\hat{\bar{v}}^0 = 48.18$ mph. So, there is a 15 percent drop off in space-mean speed (i.e., $\gamma_{max} = 1.149$). The basic Boltzmann type of model with $\beta = 0.0861$ yields poor predictions of the speed-density. It gives a d^2 of 0.0495, which is outside of our criterion (0.003) by more than an order of magnitude. The generalized Boltzmann type of model with $\beta = 0.091$ and $\lambda = 0.622$ gives a d^2 of 0.0467, which does not show significant improvement. These values of β and λ were the best possible fit values (the ones that gave the least d^2) one could determine by trying all the possible pairs of β and λ. The failure of the Boltzmann type of models at this density has been reported previously (4) and is partly due to the large pileup of cars around v^0_{min}. Figures 5 and 6 show the speed-density obtained from the basic and generalized Boltzmann models respectively.

This problem of pileup of vehicles at low speeds is relieved by the introduction of the shift parameter γ in the modified Boltzmann type of model (Fig. 7). The best fit values of γ and β, 1.116 and 0.036 respectively, for the modified basic Boltzmann model gives a d^2 of only 0.0012, which is well within our criterion. Thus, the introduction of a shifting process has resulted in a reduction of d^2 by a factor of 40. Data given in Table 1 further show that the results of using the shift model ($\beta = 0$, $\gamma = \gamma_{max} = \hat{\bar{v}}^0/\hat{\bar{v}} = 1.149$) give a d^2 of 0.0015, which is only slightly less optimal (and still well within the criterion) than the best possible γ and β fit case for the modified basic Boltzmann model (Fig. 8).

CONCLUSIONS AND APPLICATIONS

The results of this study have shown that the Boltzmann type of statistical models put forward by Prigogine et al. can be improved significantly by the incorporation of a shifting process, which results in a modified Boltzmann type of model when the shifting process is included in the basic Boltzmann model. A shift model that uses this shifting process alone has also been studied and has shown good results. The important observations of the 2 new models, as well as the basic and generalized Boltzmann type of models, are summarized below.

1. The basic Boltzmann type of model is described by the interaction and relaxation processes. The main objective of this model and of the other 3 models is to predict speed-density function at any vehicle concentration based on the knowledge of the desired speed-density function. This model says

$$f(v) = [f^0(v)]/[1 + \beta(v - \bar{v})]$$

It can be shown that there is only one value of β that normalizes f (6) for given knowledge of $f^0(v)$ and the average speed $\bar{v}$ at that concentration. Therefore, given the $\bar{v}$ - c relation and the knowledge of $f^0(v)$, one can easily compute the speed-density function for any concentration.

It can also be shown that there is only one value of average speed $\bar{v}$ that normalizes f (6) for given knowledge of $f^0(v)$ and the value of β at that concentration. Therefore, given the β - c relation and the knowledge of $f^0(v)$, one can easily compute the $\bar{v}$ - c relation for the freeway and its corresponding speed-density functions for every concentration. It should be mentioned here that there is no need to know P and T [$\beta = (1 - P)T$] in the computations given above. However, the fact that this model gives poor validation results indicates that it has little application value.

2. In the generalized Boltzmann type of model,

$$f(v) = [f^0(v) + \lambda\beta\delta(v - \bar{v})]/[1 + \lambda\beta + \beta(v - \bar{v})]$$

an adjustment process is added to the interaction and relaxation processes. In this case, 2 parameters, λ and β, are to be determined. However, β can no longer be determined by merely normalizing f(v). A direct way to obtain β is to estimate P and T as functions of concentration, for which no relations are available at present, and then to obtain λ by normalization. An indirect way is to use (4)

$$\beta = (\bar{v}^0 - \bar{v})/\sigma_v^2$$

and determine λ by normalization. In this case, we need the relation of σ_v^2 with traffic concentration.

Therefore, given the β - c (or σ_v^2 - c) and $\bar{v}$ - c relations and the knowledge of $f^0(v)$, one can compute the speed-density function for any concentration by selecting a value of λ that meets the normalization requirement of the speed-density function. Or, given the β - c and λ - c relations and the knowledge of $f^0(v)$, one can compute the $\bar{v}$ - c relation for the freeway and its corresponding speed-density function for any concentration. However, at present there is no easy way to determine the β - c or λ - c relation for a freeway. In addition to these difficulties, the validation results show that the generalized Boltzmann type of model produces poor results for all possible combinations of β and λ. Therefore, we again argue that this model is of little importance in terms of application.

3. The modified Boltzmann type of model was obtained by replacing the desired speed-density in the relaxation process of the basic Boltzmann type of model by a relaxed speed-density, produced by the shifting process. This results in

$$f(v) = [\gamma f^0(\gamma v)]/[1 + \beta(v - \bar{v})]$$

The shift parameter γ is concentration-dependent. Although β is determined if P and T are determined, it is not independent of γ because, when the value of γ is assigned, there is only one β that normalizes f(v).

Furthermore, our validation results suggested that the shifting process is far more important than the other processes. Therefore, we should determine γ first and obtain β by normalizing f(v). The validations show good results for the best possible value of γ, indicating a potential application of the modified Boltzmann type of model. To make use of this model, however, requires that a direct relation between γ and concentration be determined.

Therefore, given the γ - c and $\bar{v}$ - c relations and the knowledge of $f^0(v)$, one can compute the speed-density function for any concentration by selecting a value of β that meets the normalization requirement of the speed-density function. Or, given the γ - c and β - c relations and the knowledge of $f^0(v)$, one can compute the $\bar{v}$ - c relation for the freeway and its corresponding speed-density function for any concentration. However, at present there is no easy way to determine the γ - c and β - c relations for a freeway, and the resulting calculations will require the need of a computer program to compute f(v) by meeting the normalization requirements of the density function.

4. The shift model is obtained by using the shifting process alone and neglecting the interaction, relaxation, and adjustment processes. The resulting model

$$f(v) = \gamma_{max} f^0(\gamma_{max} v) = (\bar{v}^0/\bar{v})\, f^0[(\bar{v}^0/\bar{v})\, v]$$

was shown in the validation to be only slightly less optimal than the best possible modified Boltzmann type of model and still much better than the basic and generalized Boltz-

mann type of models. This suggests that the concept of the shifting process by itself plays a much more dominant role to describe traffic than when it is compared with the interaction, relaxation, and adjustment processes in the Boltzmann type of models. The shift model has great application value, for we need only to know—in addition to the desired speed-density—the mean speed at any vehicle concentration ($\overline{v}$ - c relation).

Therefore, given the $\overline{v}$ - c relation (or, equivalently, γ_{max} - c relation, because $\gamma_{max} = \overline{v}^0/\overline{v}$) and the knowledge of $f^0(v)$, one can compute the speed-density function f(v) for any concentration. The relation of $\overline{v}$ - c and $f^0(v)$ for any freeway can be obtained relatively with less difficulty as compared to the γ - c or β - c relation.

Many theories have been developed to obtain the $\overline{v}$ - c relation. For example, Pipes (7) suggests

$$\overline{v} = \overline{v}^0[1 - (c/c_j)]^n, \quad n > 0 \tag{19}$$

Using $\gamma_{max} = \overline{v}^0/\overline{v}$, γ_{max} can be plotted easily. Figure 9 shows the γ_{max} - c relation based on our experimental data. If $\gamma_{max}(c)$ is known , the shift model is able to predict the speed-density at any concentration through

$$f(v) = \gamma_{max} f^0(\gamma_{max} v) = [\overline{v}^0/\overline{v}(c)]\, f^0\{[\overline{v}^0/\overline{v}(c)]\, v\} \tag{20}$$

If the $\overline{v}$ - c relation is that specified in Eq. 20, then

$$\gamma_{max} = [1 - (c/c_j)]^{-n}, \quad n > 0 \tag{21}$$

Clearly, $\gamma_{max} \geq 1$, $\gamma_{max} = 1$ when $c = 0$, and $\gamma_{max} = \infty$ when $c = c_j$.

The space-speed-density is definitely related to space-headway distributions and, consequently, influences the lane-change and car-following behavior. We expect that the more accurate speed-density prediction will eventually help to develop better freeway control strategy.

REFERENCES

1. Munjal, P. K., and Pahl, J. An Analysis of the Boltzmann-Type Statistical Models for Multi-Lane Traffic Flow. Transportation Research, Vol. 3, 1969, pp. 151-163.
2. Prigogine, I., and Andrews, F. C. A Boltzmann-Like Approach for Traffic Flow. Op. Res., Vol. 8, 1960, pp. 789-797.
3. Prigogine, I., Resibois, P., Herman, R., and Anderson R. L. On a Generalized Boltzmann-Like Approach for Traffic Flow. Bull. Acad. r. Belg. Cl. Sci., Vol. 48, 1962, pp. 805-814.
4. Gafarian, A. V., Munjal, P. K., and Pahl, J. An Experimental Validation of Two Boltzmann-Type Statistical Models for Multi-Lane Traffic Flow. Transportation Research, Vol. 5, 1971, pp. 211-224.
5. Breiman, L. Space-Time Relationships in One-Way Traffic Flow. Transportation Research, Vol. 3, 1969, pp. 365-376.
6. Munjal, P. K., Hsu, Y. S., and Carpenter, R. Experimental Validation of Modified Boltzmann-Type and Shift Models for Multi-Lane Traffic Flow. System Development Corp., Rept. TM-4638/007/01, 1971.
7. Pipes, L. A. Car Following Models and the Fundamental Diagram of Road Traffic. Transportation Research, Vol. 1, 1967, pp. 21-29.

APPENDIX

BASIC AND GENERALIZED BOLTZMANN TYPE OF MODELS

The speed-density function f(x, v, t) was assumed by Prigogine et al. (2, 3) to change with time only as a result of 3 traffic processes: relaxation, interaction, and adjustment; i.e.,

$$df/dt = (\partial f/\partial t) + v(\partial f/\partial t) = (\partial f/\partial t)_{rel} + (\partial f/\partial t)_{int} + (\partial f/\partial t)_{adj} \tag{22}$$

and

$$(\partial f/\partial t)_{rel} = -[(f - f^0)/T] \tag{23}$$

$$(\partial f/\partial t)_{int} = (1 - P)(\bar{v} - v)f \tag{24}$$

$$(\partial f/\partial t)_{adj} = \lambda(1 - P)[\delta(v - \bar{v}) - f] \tag{25}$$

where

P = probability of passing = $1 - (c/c_j)$;
T = relaxation time = $[\tau(1 - P)]/P$;
λ = unknown parameter for adjustment;
c_j = jam concentration; and
τ = proportionality constant.

The parameters T, P, and λ are all functions of concentration.

For time-independent, space-homogeneous traffic flow, the solution of Eq. 22 is

$$f[1 + \lambda\beta + \beta(v - \bar{v})] = f^0 + \lambda\beta\delta(v - \bar{v}) \tag{26}$$

where $\beta = (1 - P)T$; or

$$f(v) = [f^0(v) + \lambda\beta\delta(v - \bar{v})]/[1 + \lambda\beta + \beta(v - \bar{v})] \tag{27}$$

Equation 27 is called the generalized Boltzmann type of model for $\lambda \neq 0$ and the basic Boltzmann type of model for $\lambda = 0$.

We note that $c = 0$ implies that $P = 1$ and $T = 0$; thus, $\beta = 0$. In this case, Eq. 27 reduces to

$$f = f^0$$

When $c = c_j$, we have $P = 0$ and $T = \infty$; thus, $\beta = \infty$. In this case, Eq. 27 reduces to

$$f = \delta(v - \bar{v})$$

Both extreme cases represent physical realism. It is also interesting to see how λ and β play the role in the relation between f and f^0 for arbitrary c, where $0 < c < c_j$.

For the basic Boltzmann type of model we have $\lambda = 0$ and

$$f(v) = [f^0(v)]/[1 + \beta(v - \bar{v})] \tag{28}$$

It is easy to see that for $v > \bar{v}$, $1/[1 + \beta(v - \bar{v})] < 1$; for $v = \bar{v}$, $1/[1 + \beta(v - \bar{v})] = 1$; and for $v < \bar{v}$, $1/[1 + \beta(v - \bar{v})] > 1$. The role of β is thus to skew the speed-density

Figure 10. Average speed-concentration relation for basic and generalized Boltzmann type of models (2-lane).

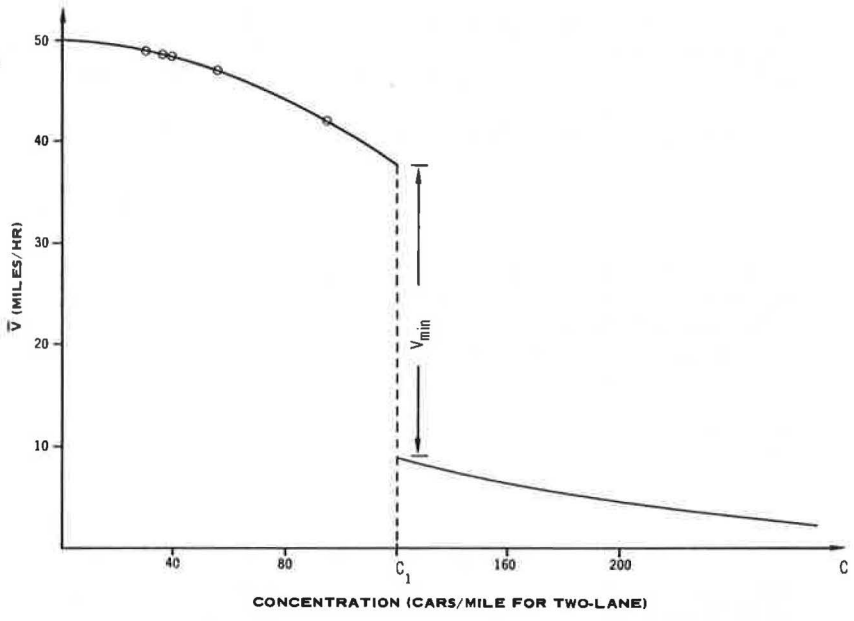
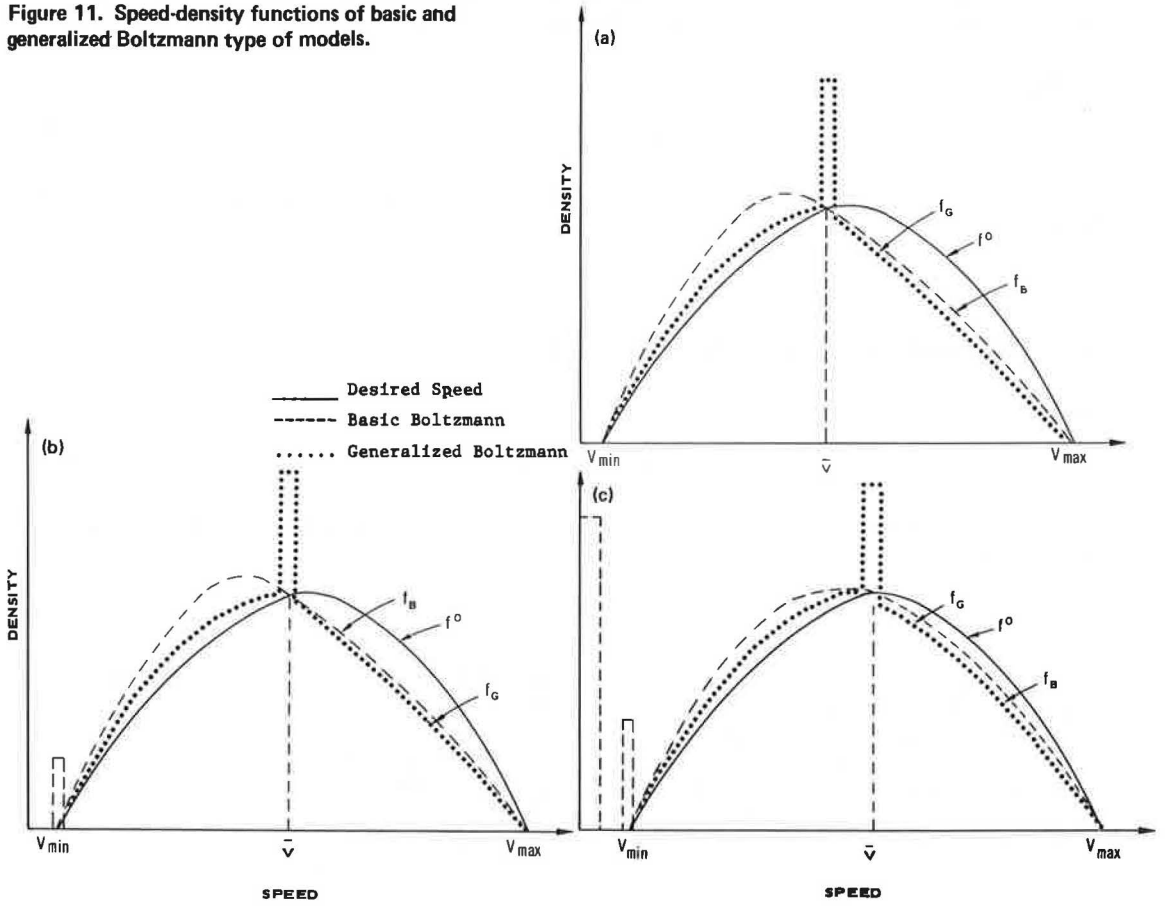


Figure 11. Speed-density functions of basic and generalized Boltzmann type of models.



toward lower values and subsequently to produce a $\bar{v}$ smaller than $\bar{v}^0$. In other words, the speed-density f decreases to a smaller value from the desired speed-density f^0 when speeds are higher than the space-mean speed of f, remains the same when equal, and increases when lower.

Furthermore, with the adjustment term $\lambda \neq 0$, we have

$$1/[1 + \beta(v - \bar{v}) + \beta\lambda] < 1/[1 + \beta(v - \bar{v})]$$

and the resulting generalized Boltzmann speed-density decreases everywhere between v_{min} and v_{max} except at $v = \bar{v}$, where there is an increase of density represented by the δ-function of magnitude $\lambda\beta/(1 + \beta\lambda)$.

The Boltzmann type of models are also influenced by vehicle concentrations. That is, we have these 3 cases discussed below.

Case A

$$1 + \lambda\beta - \beta(\bar{v} - v_{min}) > 0 \tag{29}$$

This is the most realistic case, and a previous validation study (4) concentrated on this case. Figure 10 shows the speed-concentration ($\bar{v}$ - c) curve corresponding to the Boltzmann type of models. The 5 circles are the same 5 flow levels shown in Figure 2. Case A corresponds to the range $0 \leq c < c_1$, case B corresponds to $c = c_1$, and case C corresponds to $c_1 < c \leq c_j$. The Boltzmann type of models produces a drop in $\bar{v}$ at $c = c_1$ by an amount of v_{min}. We shall explain this point in the discussion of cases B and C.

Case B

$$1 + \lambda\beta - \beta(\bar{v} - v_{min}) = 0 \tag{30}$$

This case corresponds to the situation when $c = c_1$. The average speed is determined from Eq. 30 as

$$\bar{v} = \lambda + v_{min} + (1/\beta)$$

It can be seem from Eq. 27 that f has a pole at $v = v_{min}$.

Case C

$$1 + \lambda\beta - \beta(\bar{v} - v_{min}) < 0 \tag{31}$$

This case corresponds to the situation when $c_1 < c \leq c_j$. In this case there exists a $v_1 > v_{min}$ such that

$$1 + \lambda\beta - \beta(\bar{v} - v_1) = 0$$

This means that f is singular at v_1 and negative for $v_{min} \leq v < v_1$, which is absurd from a physical point of view. To compensate for this, Prigogine et al. (3) had to assume that some vehicles were traveling at 0 speed. Comparing case C to case B shows that there is a reduction of average speed of v_{min} or a decreasing flow, $\Delta q = c_1 v_{min}$, at $c = c_1$. That is, the Boltzmann type of models produce a discontinuity in the $\bar{v}$ - c (speed-concentration) curve as shown in Figure 10. Speed-density functions of the basic and generalized Boltzmann type of models corresponding to the 3 concentration ranges, $0 \leq c < c_1$, $c = c_1$, $c_1 < c \leq c_j$ (cases A, B, and C respectively), are shown in Figure 11. Only case A is promising in terms of physical reality.

LANE-CHANGE FREQUENCIES IN FREEWAY TRAFFIC FLOW

Juergen Pahl, University of California, Los Angeles

Lane-change frequencies of exiting vehicles close to their intended off-ramp and of through vehicles were determined empirically for an 8-, a 6-, and a 4-lane freeway site. The results were obtained for various flow levels and as a function of upstream distance from the off-ramp. The data base used in this study was acquired by a methodology that involved aerial photographic techniques and that resulted in digitized space-time trajectories of all vehicles traversing the study sites. For each studied flow level at the various sites, lane-change frequencies are presented for each lane pair as well as for both exiting and through vehicles independent of lanes. The results show clearly that the greatest frequency of exiting-vehicle lane changes is toward the right lane and that a corresponding increased frequency of through-vehicle lane changes is toward the left lane. At the off-ramp, where the exiting vehicles leave the freeway, through vehicles have been shown to move back toward the right lane and to have high lane-change frequencies.

•WHEN a driver on a freeway far from his intended exit catches up to a slower driver in front of him in the same lane, he will usually attempt to change lanes in order to maintain his speed. However, if the driver intends to exit at a nearby downstream off-ramp, his behavior will be different because his primary concern will be to move toward the right lane. Such differences in lane-change characteristics of exiting-vehicle drivers in the vicinity of their intended exit from those of through-vehicle drivers may result in special operational and safety problems in the vicinity of off-ramps.

Some limited research results on lane-change behavior in freeway traffic flow are available (1, 2, 3, 4); but there is currently a strong need for more detailed research results to improve the understanding of how the lane-change characteristics affect the operational and safety features of freeway traffic flow. In response to this need, this paper presents data that detail the lane-change frequencies for exiting vehicles upstream of their intended off-ramp as well as for through vehicles as a function of flow level and distance from the off-ramp. The necessary data base for this study was acquired by the use of a recently developed aerial photographic technique (5).

AVAILABLE DATA

The 3 freeway sites selected each contained an off-ramp and an upstream section of the freeway. The sites were situated in the Los Angeles area and consisted of an 8-lane freeway (Ventura Freeway, westbound, at White Oak Avenue off-ramp), a 6-lane freeway (Santa Ana Freeway, southbound, at Washington Boulevard off-ramp), and a 4-lane freeway (Newport Freeway, northbound, at Katella Avenue off-ramp). More details as to the location and geometric design features of these freeway sites are presented in another report (5).

The data were collected with a 70-mm Maurer camera operated from a helicopter that hovered over the studied freeway site. The camera took pictures at time intervals of 1 sec, and the length of the photographed freeway section was of the order of 1 mile. The pictures were scanned in successive order so as to obtain digitized positions of vehicles. A specialized software system deduced individual trajectories from the digitized vehicle coordinates. The resulting trajectories yield the space-time history of each vehicle within the studied freeway site. They include longitudinal and lateral position and speed for each car in time intervals of 1 sec.

The acquired data were subjected to a statistical analysis to isolate those time periods in which the average flow level remained constant. The length of the obtained time periods ranged from 1.5 to 15 min. The data were aggregated into groups where the average flow levels of the selected time periods differ only within a small range in order to improve the statistical significance of the derived analysis results. Table 1 gives the resulting data groups that were used for this study (more details are given in another report, 5).

On the 8-lane freeway for medium and high flow, data in the range from 5,250 to 9,450 ft were obtained from basically the same data sample that was used to obtain the data in the range from 0 to 4,800 ft. For this purpose, the license numbers of the cars traversing the freeway site were photographed by a ground camera system and later assigned to the corresponding car trajectories. License numbers were also recorded at the first off-ramp downstream from the study site, and thus lane-change frequencies could be obtained from the data taken at the study site for cars that were between 1 and 2 miles from their intended off-ramps. In this case, through vehicles were redefined as all vehicles that did not exit at the downstream off-ramp. Therefore, the results obtained for through vehicles between 1 and 2 miles are based on partially the same vehicle sample as the results between 0 and 1 mile—a fact to be considered very carefully in the interpretation of the results.

The acquired data were also analyzed to obtain traffic flow characteristics, such as lane distributions of vehicles, overall speed distributions, speed distributions for each lane, and gap acceptance, as a function of upstream distance of through and exiting vehicles from the exit ramp.

DATA ANALYSIS

For the data analysis, each of the study sites was divided into zones 600 ft in length starting at the nose of the off-ramp and progressing upstream. In each zone all lane changes of exiting and through vehicles were determined and classified by lane-change type; i.e., the number of changes from lane 1 to lane 2, from lane 2 to lane 1, from lane 2 to lane 3, and so on was obtained for the various data groups. Trajectories that entered the sites but terminated before reaching the off-ramp were excluded from this analysis because they could not reveal whether they were exiting- or through-vehicle trajectories (such trajectories can be generated at the end of a data collection period).

If the number of changes by through vehicles from lane i to lane j in the zone at distance d from the off-ramp is denoted by $n_{i,j}(d)$, the frequencies of through-vehicle lane changes

$$L_{T,i,j}(d) = n_{i,j}(d)/[N(d) \times 0.1135 \text{ mile}] \quad (1)$$

were formed in each zone. N(d) is the total number of through-vehicle trajectories that were available in the zone, whereby a trajectory that did not extend all the way through the zone was only counted as that fraction of one that equaled the available trajectory length in the zone divided by the length of the zone. The division by 0.1135

Table 1. Average flow levels of study sites.

Number of Freeway Lanes	Flow	Avg Number of Vehicles/Hour		Time Period (sec)
		Through	Exiting	
8	Low	4,450	437	1,040
	Medium	5,520	595	1,120
	High	7,320	771	2,525
6	Medium	4,440	484	400
	High	5,680	560	180
4	Medium	2,520	427	1,405

mile, the length of a zone, in Eq. 1 yields results in terms of number of through-vehicle lane changes per through-vehicle-mile in each zone.

Correspondingly, the frequencies of exiting-vehicle lane changes,

$$L_{E,i,j}(d) = e_{i,j}(d)/[E(d) \times 0.1135 \text{ mile}] \quad (2)$$

were formed in each zone, where $e_{i,j}(d)$ is the number of exiting-vehicle lane changes from lane i to lane j in the zone at distance d from the off-ramp, and E(d) is the total number of exiting-vehicle trajectories [analogous to N(d)] that were available in the zone. The results from Eq. 2 yield the number of exiting-vehicle lane changes per exiting-vehicle-mile in each zone.

Furthermore, the frequency of any lane change of a through vehicle per mile was computed by forming the sum $\sum_{i,j} L_{T,i,j}(d)$. Correspondingly, the frequency of any exiting-vehicle lane change per mile was computed by forming the sum $\sum_{i,j} L_{E,i,j}(d)$.

RESULTS

The obtained lane-change frequencies per vehicle-mile are shown in Figures 1 through 6. For the sake of simplicity, all data points are plotted at the midpoints of the zones for which they were obtained, although they actually present an average value over the entire zone.

The sample sizes represented in these figures range from a few to about a hundred lane changes, depending on the data group (the actual results are presented in an earlier report, 5, Appendix F). This means that many of the variations of the plotted data points are statistically insignificant. Therefore, smooth curves have been hand-fitted through the data points with an attempt to indicate only statistically significant features of the data. In the figures showing lane-change frequencies classified by lane-change types, data points for very small sample sizes were omitted.

SUMMARY AND CONCLUSIONS

The largest sample sizes occur in the 3 data groups for the 8-lane freeway. For each of these data groups, Figures 4, 5, and 6 show a pronounced maximum of through-vehicle lane changes from the left lanes toward the right lanes in the zone right at the off-ramp, i.e., from 0 to 600 ft upstream of the off-ramp nose. In this zone, the lane-change frequencies from lanes 2 to 1 are consistently higher than the corresponding frequencies of lane changes from lanes 3 to 2 and lanes 4 to 3. Another set of maxima appears in the zones from 3,600 to 4,800 ft but for the opposite types of lane changes, namely, lane changes from lanes 1 to 2, lanes 2 to 3, and lanes 3 to 4.

These features can be explained if seen in relation to the corresponding behavior of exiting vehicles shown also in Figures 4, 5, and 6. Pronounced maxima of exiting-vehicle lane changes from lanes 2 to 1 occur in the zone from 1,800 to 2,400 ft for low and medium flow and in the zone from 2,400 to 3,600 ft for high flow of the 8-lane freeway site. Maxima for lane changes from lanes 3 to 2 and lanes 4 to 3 are less pronounced and appear at larger distances from the off-ramp than the maxima for the exiting-vehicle lane changes from lanes 2 to 1.

Combining these features for through and exiting vehicles reveals that through vehicles change lanes toward the left as exiting vehicles change lanes toward the right upstream of the off-ramp. As the exiting vehicles exit in the zone from 0 to 600 ft, through vehicles move back into the right lane with a high lane-change frequency; this also causes increased lane-change frequencies of through vehicles from lanes 3 to 2 and lanes 4 to 3.

Analogous behavior, although less pronounced, is found for the 6-lane freeway site for medium flow. For high flow on the 6-lane freeway site, exiting-vehicle lane changes were too sparse to be shown on a meaningful graph. But data shown in Figure 2 suggest that the frequencies of through-vehicle lane changes from lanes 2 to 1 and lanes 3 to 2 do not have maxima in the zone from 0 to 600 ft from the off-ramp nose.

Figure 1. Four-lane site, medium flow.

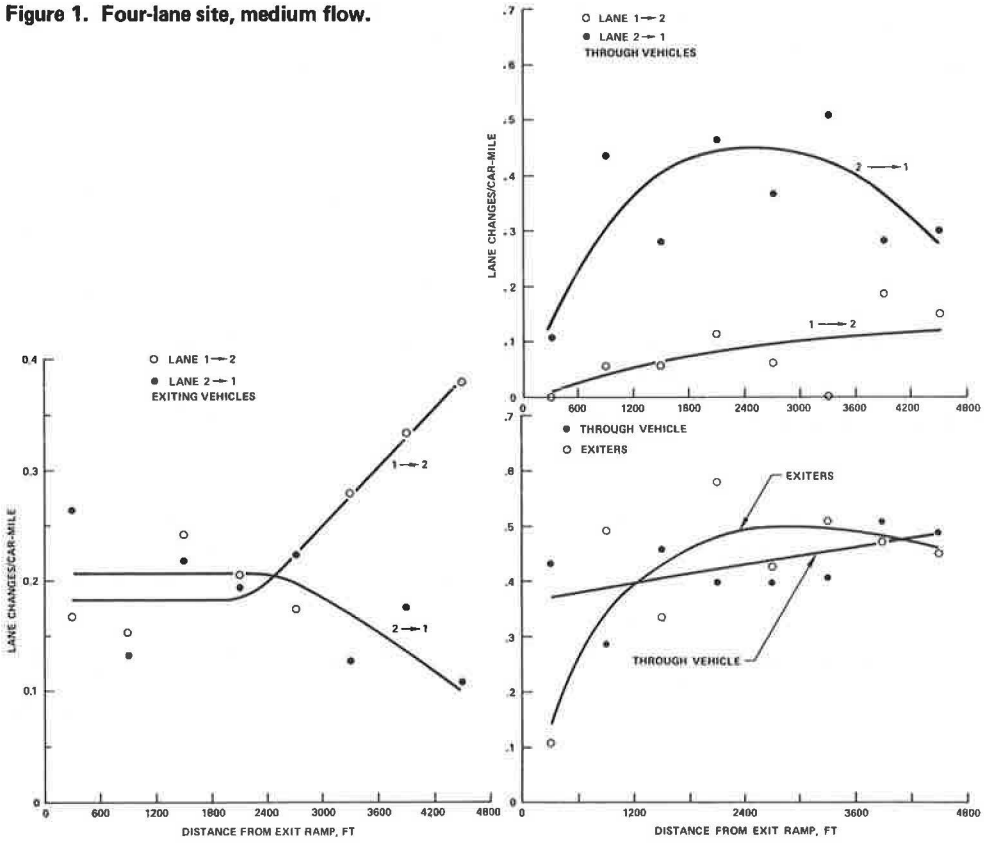


Figure 2. Six-lane site, high flow.

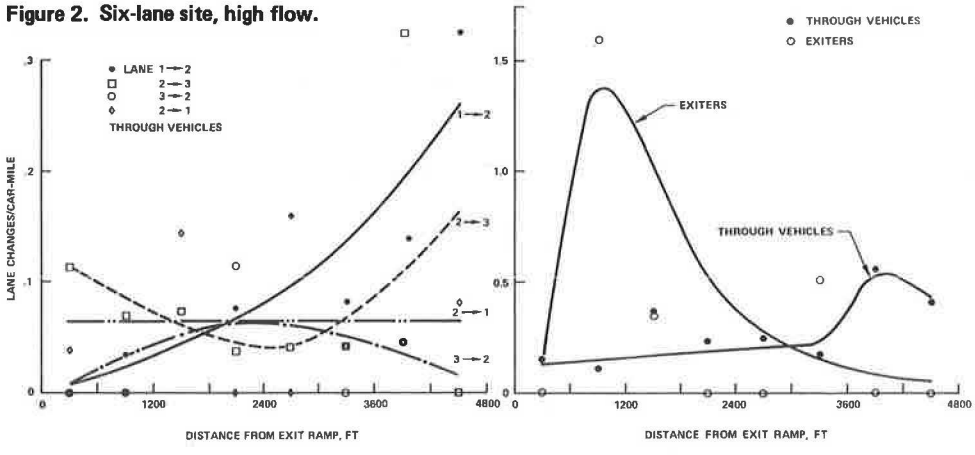


Figure 3. Six-lane site, medium flow.

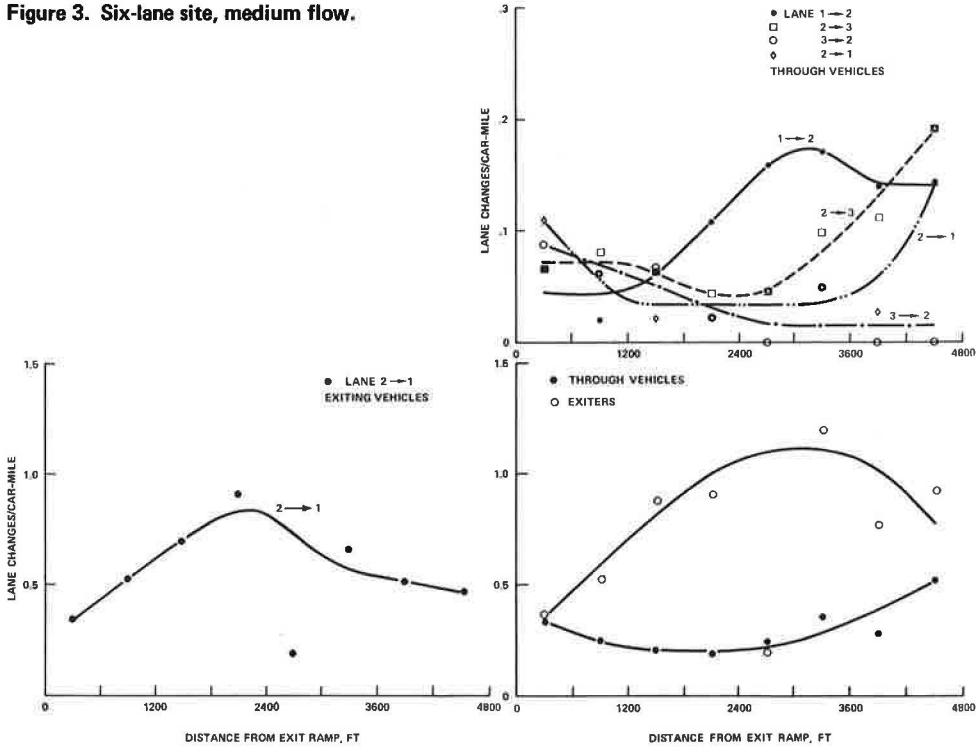


Figure 4. Eight-lane site, high flow.

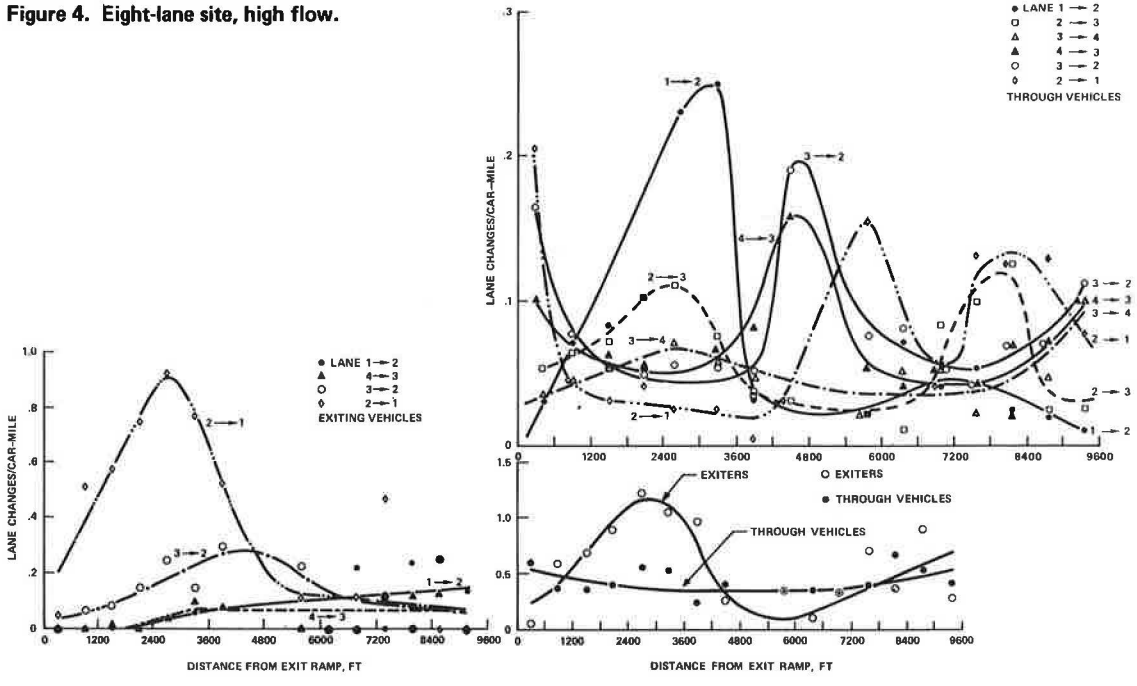


Figure 5. Eight-lane site, medium flow.

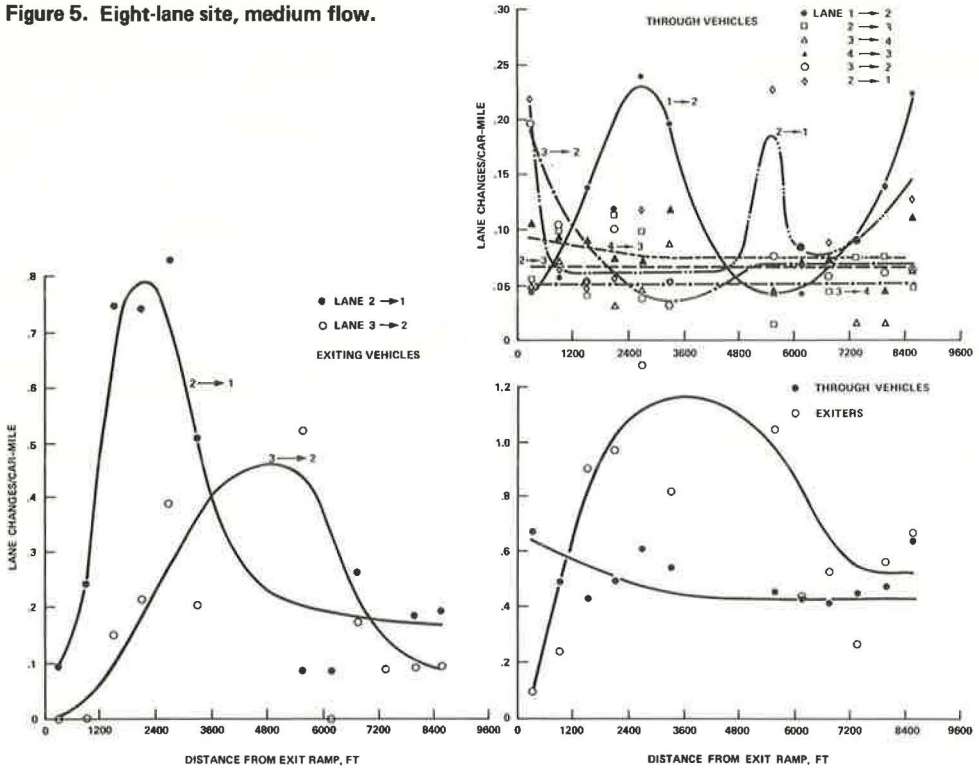
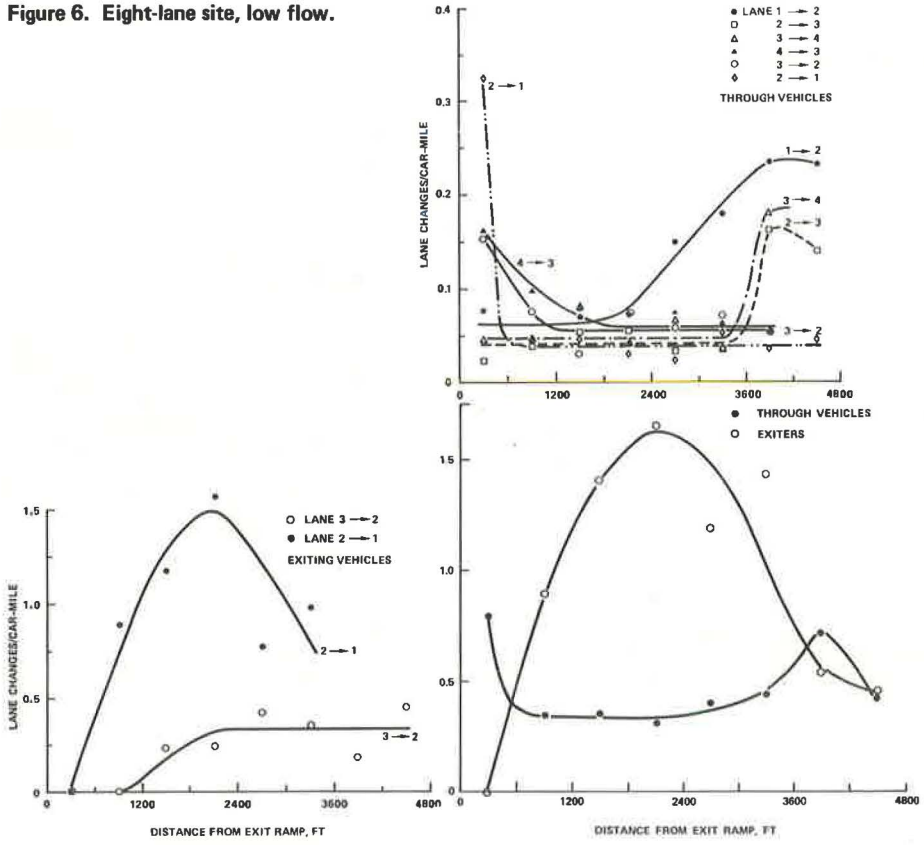


Figure 6. Eight-lane site, low flow.



For the 4-lane freeway site, Figure 1 shows that the maximum of the frequency of exiting-vehicle lane changes from lanes 2 to 1 occurs in the zone from about 1,200 to 3,600 ft from the off-ramp nose. Through vehicles exhibit a corresponding increase in lane changes from lanes 1 to 2 in the same zone. No maximum occurs for through-vehicle lane changes from lanes 2 to 1 in the zone from 0 to 600 ft from the off-ramp nose.

REFERENCES

1. Worrall, R. D., and Bullen, A. G. R. An Empirical Analysis of Lane Changing on Multilane Highways. Highway Research Record 303, 1970, pp. 30-43.
2. Oliver, R. M., and Lam, T. Statistical Experiments With a Two-Lane Flow Model. Proc. Third Internat. Symposium on Theory of Traffic Flow, New York, 1967, pp. 170-180.
3. Clinton, J. W. Lane Changes on an Urban Freeway. John C. Lodge Freeway Traffic Surveillance and Control Research Project, Detroit, 1962.
4. Wynn, F. H. Weaving Practices on One-Way Highways. Bureau of Highway Traffic, Yale Univ., May 1946, pp. 9-57.
5. Pahl, J., Charles, S. E., and Knobel, H. C. Exit Ramp Effects on Freeway System Operation and Control. Univ. of California, Los Angeles, Eng. Rept., 71-49, 1971.

DISCUSSION

R. D. Worrall, Peat, Marwick, Mitchell and Company

The author presents the results of an extensive, empirical analysis of lane changing on freeways, based on a set of photographic data collected at a series of 4-, 6- and 8-lane freeway locations in California. The data set on which his analysis is based represents an extremely rich data source for the analyst of freeway traffic flow; it is perhaps the richest source of its type that has been developed to date.

Attention is focused primarily on variations in lane-changing frequencies, expressed as a function of distance from the downstream off-ramp. Separate graphical analyses are presented both for through- and for exiting-vehicle behavior, and also for varying levels of total traffic flow. The results indicate a clear pattern of compensatory behavior on the part of through vehicles in response to the lane-changing maneuvers of exiting vehicles. They show that through vehicles tend, in the main, to move away from the lane adjacent to the off-ramp at a considerable distance upstream from the ramp nose and to return to that lane as soon as the exiting traffic has departed from the through pavement.

In commenting on Pahl's paper, I would like to focus my remarks on 2 topics: The first of these deals with a collection of essentially methodological issues; the second is concerned with the potential implications of research of this type for the design or traffic-control engineer.

Previous empirical research on lane changing (1, though, as the author notes, it is somewhat sparse in its content) has tended to suggest strongly that lane changing is best treated as a stochastic phenomenon, subject only to relatively limited systematic influences. One such influence is clearly the presence of an exit ramp. In this context, the somewhat "noisy" statistical pattern that emerges from the present data set is perhaps not surprising. It would, however, be extremely interesting if the simple, graphical analyses presented here could be supplemented by a more rigorous statistical examination of the data. The writer appreciates that this is by no means a trivial undertaking, and he fully recognizes the author's rationale for presenting simple, "hand-fitted" curves in his paper. It may, however, be argued with some validity that such a simple portrayal may in fact lead to deceptive, simple, and potentially erroneous conclusions.

One might, for example, find it useful to examine the variability in average lane-changing frequencies over a succession of short time periods for each given location and each flow level. The result of this analysis may then be compared in turn with the aggregate pattern of behavior presented here as a function of distance upstream from an exit ramp. Previous research conducted by the writer (1) suggests that such location-specific variance in lane-changing behavior may in fact be higher than that observed between different locations.

In a similar vein, the same research indicated only a relatively weak tendency for overall minute-by-minute lane-changing intensities to vary systematically with respect to ramp location and design. The distribution of maneuvers at each location was essentially random and the major impact of adjacent ramp terminals was reflected in the presence of intermittent platoons of vehicles entering the freeway from an on-ramp. It is also interesting to note here that analysis (1) of lane-changing behavior in the vicinity of one cloverleaf interchange in Chicago yielded a profile as shown in Figure 7. The lane-change rate gradually increased on the approach to the interchange, decreased within the interchange area, and suddenly increased immediately downstream of the entrance ramp.

Similar statistical comparisons may also be made of the observed patterns of behavior in a variety of different ways. One possible mechanism is to describe the overall pattern of lane-changing behavior at a given location and a given flow level in terms of a simple "lane-change transition matrix." Figure 8 shows 6 such matrices, based on data developed at Northwestern University (6). Estimation of a set of compatible matrices of this type, for each of a series of different locations or different flow levels or both, can provide the analyst with a simple base for the statistical comparison of varying patterns of lane-changing behavior both at a given location and across several locations or flow levels or both. A variety of other alternative descriptors may equally well be suggested, each of which might perhaps provide a useful statistical supplement to the graphical analysis presented here.

Of perhaps more important practical concern than the preceding, somewhat academic comment on methodology is the question of the manner in which the results of the present study may be put to use by the freeway designer or traffic operations engineer. I would like to address my final remarks to this point.

Clearly, the analytical results presented here, and particularly the level of detail incorporated within the underlying data set, represent a significant resource for the traffic flow-theoretician. They also, however, have no less significant implications for the highway designer or traffic engineer.

The sympathetic relation, for example, observed between the behavior of through vehicles and that of exiting vehicles upstream of an off-ramp has clear implications both for the development of freeway control logic and also, particularly, for the identification of optimum detection and control locations for a given section of highway. Equally, the patterns of lateral turbulence observed on the approaches to an exit ramp suggest both potential constraints on throughput capacity and possible limitations on the spacing and configuration of adjacent ramp terminals. Similar implications may also, at least potentially, be drawn from the results in terms of the location and design of directional signing, the details of ramp geometrics, and the vertical and horizontal alignment of the through pavement.

Figure 7. Typical lane-changing transition matrices.

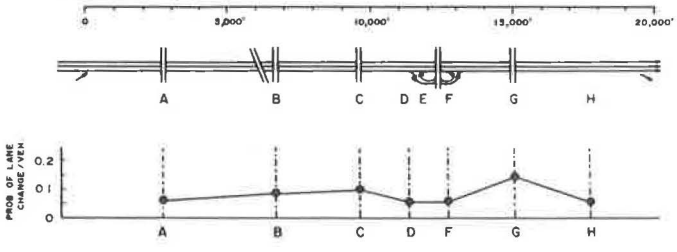
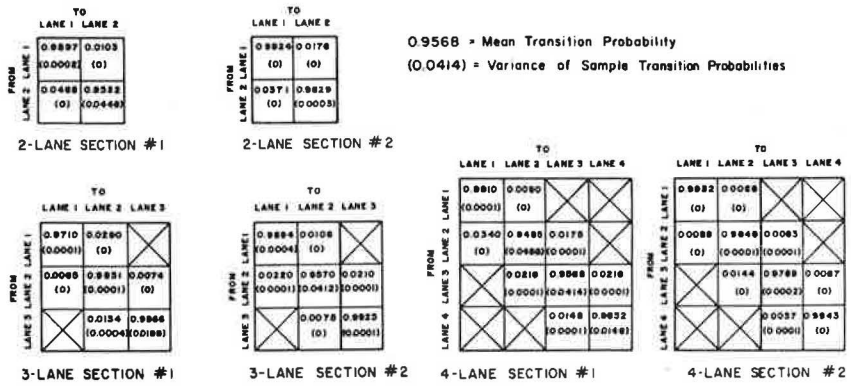


Figure 8. Variation in lane-changing frequency with distance from cloverleaf interchange on 3-lane roadway.



Clearly, examination of such questions in the context of highway design and operational control lies outside the scope of the present study. To answer them, one would need to structure at least a psuedo-experimental design that would examine the variation in lane-changing behavior at a broader mix of locations. The behavior would be expressed explicitly as a function of varying ramp configuration, sign layout, or the vertical-horizontal alignment of the freeway. The present study, though it does not address such questions, provides a valuable and substantive foundation for their analysis.

References

6. Worrall, R. D., Bullen, A. G. R., and Gur, Y. An Elementary Stochastic Model of Lane-Changing on a Multilane Highway. Highway Research Record 308, 1970, pp. 1-12.
7. Bullen, A. G. R., and Worrall, R. D. Formulating a Model of Multilane Traffic Flow. Highway Research Record 334, 1970, pp. 34-38.

COMPARISON OF TRAFFIC CIRCULATION ALTERNATIVES THROUGH SIGOP

T. E. Flynn and C. M. Siu, Wilbur Smith and Associates

Results of analyzing 2 traffic patterns—an existing roadway and traffic signal operation and a proposed traffic flow pattern with optimized signal operation—are presented. Network alternative comparisons are shown for vehicle stop and delay costs, average speeds, and travel conditions. Suggested modifications to the original SIGOP program are also presented. The purpose of the paper is to evaluate the practicability of using SIGOP results as a means of analyzing traffic circulation patterns and traffic signal timing alternatives. Conclusions are that, with minor program modifications required for each network having special traffic characteristics, SIGOP provides the practical results necessary to improve traffic conditions on a grid street system.

•THE COMPUTER continues to play an increasingly important role in the field of transportation technology. SIGOP (traffic signal optimization program), a relatively new application of the computer, assists the traffic engineer in area-wide traffic system analyses (1). The program calculates phase splits and offsets for a given set of input data such that the minimum vehicle stop and delay costs are determined for a grid street network controlled by an interconnected signal system.

This paper explains the authors' experience in applying the program to actual city networks to improve existing traffic circulation efficiency. The principal purpose of using SIGOP in the example cited here, however, was to compare alternative networks—the existing versus a proposed modified street operation system for a moderate-sized East Coast city. This analysis relied heavily on comparative vehicle stop and delay costs and time-space relations in ascertaining the practicability of implementing a "model" signal-timing program.

EXISTING ROADWAY AND TRAFFIC CHARACTERISTICS OF MODEL NETWORK

A portion of the road network of the study area now comprises a rectangular grid system with typical city block dimensions of about 400 by 600 ft; Figure 1 shows the existing road network with the proposed street operation. Major intracity traffic volumes occur on the 5 east-west arterials (Avenues A through E); heavier volumes occur on Avenues A and B. Traffic signals on these main thoroughfares operate in simultaneous progression (0 offsets); arterial movements are allocated approximately 75 sec of green time of a 120-sec cycle. Travel speeds during peak traffic conditions average approximately 20 mph, but delay is common on the north-south cross streets (Streets 1 through 10) because of a lack of signal interconnection and the long waiting period between green signal indications.

A 4- by 14-block section of the model city central business district analyzed contains mostly 2-way streets with 66 signalized intersections. The proposed traffic circulation pattern would have an extensive 1-way street network, as follows:

Direction	Existing Network (miles)	Proposed Network (miles)
1 way	1.6	6.8
2 way	8.5	3.3
Total	10.1	10.1

Avenues C and D would function as a 1-way pair, and all but one north-south cross street would operate 1-way.

SIGOP PROGRAM

An IBM 360 computer was used to run the SIGOP program. However, most computers with a FORTRAN compiler and a main core storage capacity of 300,000 bytes (or 150,000 bytes using the "overlaying" technique) can be used.

The SIGOP package consists of 6 programs—INPUTS, PHASES, OFFSET, OPTMIZ, VALUAT, and OUTPUT. After the input data are verified, phase splits for each intersection are determined based on traffic volumes. Ideal difference in offsets for each network link, based on the time needed to clear the traffic queue, is computed. A traffic simulation is then utilized to evaluate the calculated phase splits and offsets for a network that minimizes delay and stops. The following specific data are included in the output:

1. Cost estimates as a function of vehicle stops and delay;
2. Cycle splits at individual signalized intersections;
3. Intersection offsets of the optimized network; and
4. Time-space diagrams for specified study sections of road.

The SIGOP program analyzed the existing and future networks. A future traffic assignment was made based on land use, roadway characteristics, and origin-destination patterns.

Detailed intersection traffic and roadway inventory characteristics were provided in the analysis. Existing input data included

1. Cycle length, offsets (for existing conditions only), and number of phases;
2. Specific signal phase intervals, such as a delayed beginning of green, advanced beginning of red, all pedestrian time, green time, and amber time;
3. Effective number of approach and discharge traffic lanes;
4. Traffic volumes for all through and turning movements; and
5. Traffic characteristic parameters, such as proportion of trucks, degree of vehicle platooning, and vehicle speeds.

OPERATIONAL MODIFICATIONS TO SIGOP PACKAGE

Several modifications were made to the original SIGOP program so that it would be more responsive to particular conditions of the specific network analyzed. The more pertinent existing program modifications made or proposed in future computer runs based on actual experience of application are summarized below.

1. The SIGOP manual does not specifically state which of several platoons of traffic approaching an intersection should be selected as the main platoon. A policy decision was made to assign the traffic making the through movement as the main platoon approaching from the upstream intersection.
2. DLANE is an input parameter that measures the number of traffic lanes accommodating the main traffic platoon leaving an intersection. This value is used in determining DRATE, the discharge rate or the rate a queue of vehicles moves through an intersection on a green signal after having been stopped by a red signal, and CRFLO, the critical flow or the maximum observed flow per lane through an intersection for a given phase. Coding DLANE equal to ALANE, the effective number of lanes accommodating the main platoon approaching an intersection improved the determination of DRATE and CRFLO.
3. The SIGOP manual recommends the use of "artificial peripheral links" in the coding of exclusive left turns. This method has 2 disadvantages: (a) The coding procedure is complicated by the appearance of additional links, and (b) the real link loses part of its traffic volume to the artificial link. That loss affects the ideal offset and evaluation of the performance functions (vehicle stops and delays). It is proposed that traffic movements through a 3-phase intersection be "coded" as a 2-phase movement.

Figure 2 (upper portion) shows traffic movements during the signal phase that includes an exclusive left-turn subphase. This is similar in overall traffic aspects to the arrangements that are shown in the lower portion of Figure 2 and provide results just as accurate but use a simpler coding procedure. The lower signal arrangement, with 3 subphases, can be coded simply by assigning ADVAR = S to link A (where ADVAR is the advanced beginning of red) and DEBEG = S (where DEBEG is the delayed beginning of green). The absolute values of the calculated offsets will differ from those calculated by the "artificial link" method but will provide comparable results when related to adjacent intersection offsets. Phasing for the left-turn movement can be easily related to the individual intersection offsets included in the program output.

4. The link-group parameters, LKGP(3), summarize delay and stop costs for 8 different groups of intersections in the optimized network. Inasmuch as it is often desirable to analyze more than 8 separate subgroups of intersections, changes were made in the SIGOP program to allow values from 0 to 99. Ninety-nine subnetwork groups of the total network can, therefore, be analyzed separately.

5. Platoon coherence is the degree to which a group of vehicles moves at uniformly close spacing and speed along a street. The platoon coherence factor, ALPHA, is an important parameter that is often assigned a human judgment value, such as "slightly coherent" or "strongly coherent." It is suggested that ALPHA be assigned a value based on the following mathematical formula:

$$\text{ALPHA} = (\text{VOLYM}/\text{ALANE}) \times (\text{CYCLE}/3{,}600) \div (\text{GREEN}/\text{DISCH})$$

where

VOLYM = total link volume, in vehicles/hour;
ALANE = effective number of approach lanes;
CYCLE = cycle length, in sec;
GREEN = green time, in sec; and
DISCH = minimum average headway of automobiles at one intersection lane, in sec.

In general, platoon coherence is directly related to factors such as traffic volume and roadway capacity; therefore, determining platoon coherence for many street sections is more practical by the recommended formula than by observation in the field.

6. The SIGOP program treats left-turn and right-turn traffic the same with respect to its effect on increasing computer-calculated effective traffic volumes. Similar treatment for left and right turns reduces analysis accuracy because existence of left-turn traffic in a lane will hinder traffic movements over a roadway. A more realistic left-turn traffic hindrance value, LTTHE can be developed by modifying the equivalence factor, EQFAC—a factor that converts commercial or turning vehicles into an equivalent volume of passenger vehicles moving straight through an intersection. The following formula, based on other input data, modifies EQFAC and eliminates the manual approximation of the turning volume hindrance effect:

$$\text{EQFAC} = (1 - \text{FTRUK} + \text{TRUCK} \times \text{FTRUK}) \times (1 - \text{FTRDN} + \text{RIGHT} \times \text{TRNFCR} + \text{LEFT} \times \text{TRNFCL})$$

where

TRNFCR = TRNFC, a turning equivalence factor usually expressed as a value between 1.00 and 2.00;
FTRDN = RIGHT + LEFT;
RIGHT = percentage of right-turn traffic to total traffic; and
LEFT = percentage of left-turn traffic to total traffic.

For 1-lane links,

$$\text{TRNFCL} = (1 + \text{ALPHA}_L) \times (1 + \text{ALPHA}_0) \times \text{TRNFC}$$

where

$ALPHA_L$ = coherence factor of the link proper, usually expressed as a value between 0.00 and 1.00; and
$ALPHA_O$ = coherence factor of the opposite link, usually expressed as a value between 0.00 and 1.00.

For multilane links,

$$TRNFCL = (1 + ALPHA_O) \times TRNFC$$

7. Some overall changes were made in the 6 programs constituting SIGOP to effect greater computer operational efficiency. For example, the INPUTS program was divided, making a total of programs in the entire package. The PHASES program was rewritten to use the green-cycle ratio extensively in computations of the signal phase split.

RESULTS OF COMPUTER RUN

Data output from the SIGOP program provided important findings concerning cycle length, efficiency of signal operation, signal progression, and travel times.

Cycle Lengths

Estimates of optimized vehicle delay and stops for the 2 most heavily traveled streets (Avenues A and B) and total optimized network are given in Table 1. A general trend can be seen because, as the cycle length decreases, delay also decreases, whereas the number of stops increases. This suggests that, with a shorter cycle length, a vehicle is stopped a shorter overall length of time but is stopped more frequently.

The SIGOP program aided in the determinination that the existing 120-sec cycle length is not so efficient as a shorter cycle length would be. The most efficient cycle lengths for Avenues A and B are 105 and 90 sec respectively, whereas the total network operates most efficiently at 60 sec. The shorter cycle length necessary to efficiently accommodate the low traffic volumes on the numerous side streets explains the 60-sec cycle length for the total network. A 90-sec cycle length was, therefore, recommended for the following reasons:

1. Traffic operation efficiency on the numerous side streets was somewhat sacrificed to provide an acceptable level of efficiency on Avenues A and B, which have peak-hour traffic volumes equaling roadway capacity; and
2. The 90-sec cycle length time-space diagrams for the principal streets provided acceptable traffic patterns.

Optimization of Signal Operation

Data output of the SIGOP program provided the basis for the conclusion that greater efficiency in traffic flow would be achieved by the optimization of the operation of the networks' traffic signals. Results of SIGOP also helped substantiate the improvement in traffic operation anticipated with the proposed 1-way street network. A sample printout of offsets for an optimized network is given in Table 2. The relative degree of improvement resulting in traffic operation is calculated by estimating the delay and number of vehicle stops at each intersection for a given time period for each alternative network. Costs of $2.50/hour of vehicle-delay and $1.00/100 vehicle stops were assigned to these parameters.

Table 3 gives estimated vehicle stop and delay costs for a 1-hour time period for 3 street and signal operating conditions. A reduction in cost from $1,780 to $1,460 should result from the optimization of the traffic signal operation of the existing system. Optimizing signal operation and implementing the proposed street operation pattern (conversion of many 2-way streets to 1-way) should further reduce total delay and stop costs to $1,160. The additional capacity resulting from 1-way operation explains this additional savings.

Figure 1. SIGOP model network.

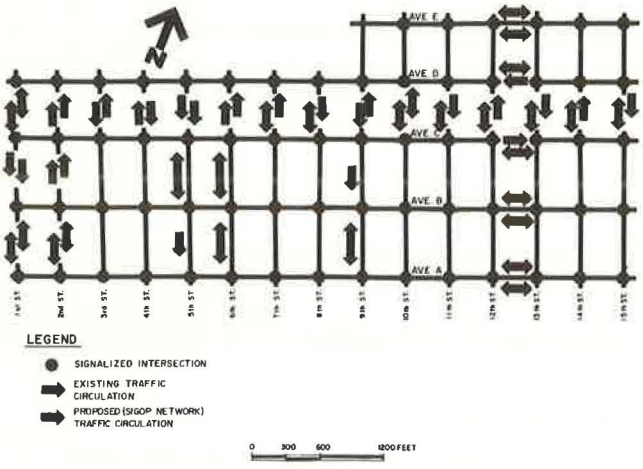


Figure 2. Analysis of exclusive left-turn lane coding.

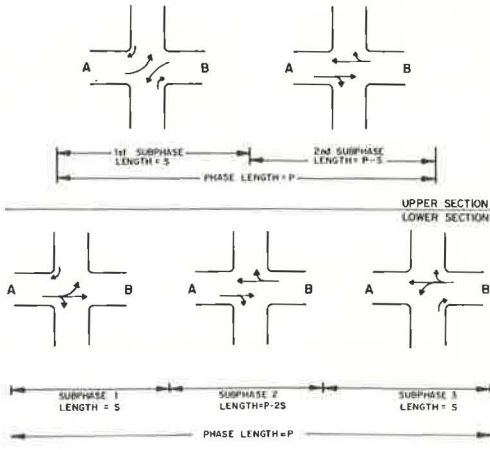


Table 1. Vehicle stop and delay estimates for different cycle lengths on optimized Avenues A and B and total network.

Cycle Length (sec)	Delay[a]			Stops[b]			Cost[c]		
	A	B	Network	A	B	Network	A	B	Network
120	31	114	371	2,950	13,280	38,230	107	418	1,310
105	31	106	334	2,620	13,450	39,270	104	400	1,228
90	30	100	304	3,560	13,040	40,480	111	380	1,165
75	31	93	263	4,160	20,140	46,170	119	434	1,119
60	36	88	241	5,460	21,340	48,700	143	433	1,089

[a]Total hourly vehicle-hours of delay due to traffic waiting at signalized intersections.
[b]Total number of hourly vehicle stops due to traffic stopping at signalized intersections.
[c]Total hourly cost due to delay ($2.50/hour of vehicle delay) and stops ($1.00/100 vehicle stops).

Table 2. Sample printout of optimized offsets, 90-sec cycle.

Intersection[a]	Offset[b]	Intersection	Offset	Intersection	Offset	Intersection	Offset	Intersection	Offset	Intersection	Offset
112	0.0	123	0.94	219	0.63	315	0.49	326	0.89	422	0.09
113	0.99	124	0.96	220	0.60	316	0.53	412	0.32	423	0.16
114	0.98	125	0.97	221	0.60	317	0.59	413	0.41	424	0.26
115	0.98	126	0.99	222	0.60	318	0.60	414	0.45	425	0.36
116	0.97	212	0.60	223	0.60	319	0.58	415	0.48	426	0.47
117	0.96	213	0.57	224	0.59	320	0.54	416	0.54	512	0.64
118	0.95	214	0.61	225	0.60	321	0.41	417	0.58	513	0.74
119	0.94	215	0.57	226	0.62	322	0.40	418	0.64	514	0.72
120	0.95	216	0.60	312	0.54	323	0.33	419	0.70	515	0.85
121	0.94	217	0.61	313	0.52	324	0.19	420	0.79	516	0.71
122	0.94	218	0.58	314	0.51	325	0.05	421	0.90	517	0.58

[a]Avenues are numbered serially as follows: A, 100's; B, 200's; C, 300's; D, 400's; and E, 500's.
[b]Optimized offsets expressed as a percentage of cycle length.

Table 3. Stop, delay, and cost estimates for street operation alternatives.

Street Operation	Traffic Signal Operation	Cycle Length (sec)	Stops	Delay	Cost
Existing	Existing	120	49,810	510	1,780
Existing	Optimized	120	40,870	420	1,460
Proposed 1-way	Optimized	90	40,480	300	1,160

Figure 3. Existing timing relations on Avenue D.

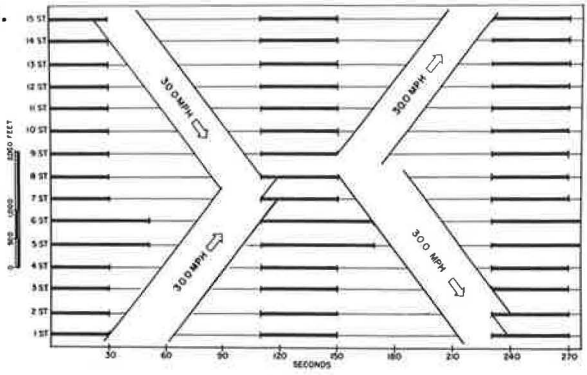


Figure 4. Proposed timing relations on Avenue D.

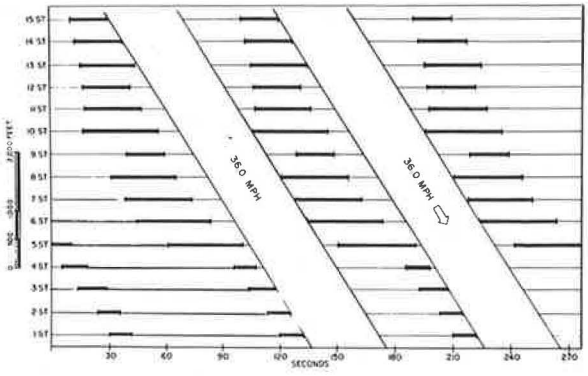


Figure 5. Existing and proposed timing relations on Fifth Street.

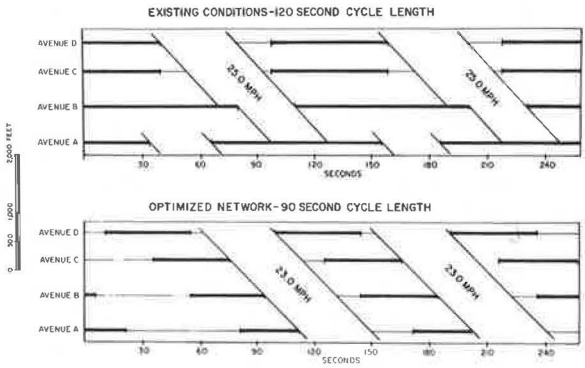
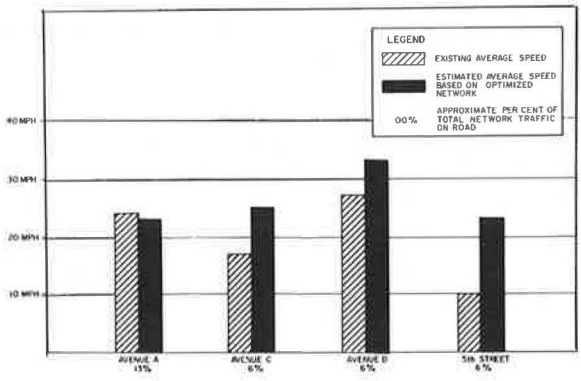


Figure 6. Existing and proposed average speed estimates.



Implementation of the proposed traffic circulation patterns, plus optimization of traffic circulation operation, should result in a 33 percent reduction in vehicle stop and delay costs. Eighteen percent of the savings is attributed to conversion to 1-way operation of many streets. A disadvantage of 1-way operation is that an increase in total vehicle-miles traveled of approximately 6 percent should result. However, the increase in average vehicle speeds on the proposed traffic network should result in an overall reduction of approximately 7 percent in total network travel time.

Signal Progression

Cross-street progression should be improved by the utilization of a general progression (versus a perfect linear progression) on the east-west arterials and by the development of cycle splits and progression responsive to cross-street traffic as well as the heavier arterial traffic. SIGOP printouts of time-space diagrams provided a convenient means of determining whether the most efficient signal operation calculated also produced realistic traffic-flow patterns that would improve the quality of driving.

Figures 3 and 4 depict the time-space relations for an existing 2-way arterial and its proposed 1-way operation respectively. One-way operation not only results in an increase in roadway capacity of about 25 percent but also provides a satisfactory progression that is currently not available.

Figure 5 shows existing and proposed time-space relations of a major cross street. Improved traffic flow in this street is reflected in the proposed 23-mph progression resulting after optimization of the network's signal operation.

Various output results were developed by varying different input parameters weighing the importance of vehicle delay to stops and of critical flow to total flow. Although progression on Avenues A and B remained relatively constant for the different parameters used in the input data, there were substanial differences on the less traveled Avenues C and D. The most efficient network did not achieve satisfactory progression on Avenues C and D. A less efficient network with adequate progression on those 2 arterials was selected as the system to be recommended for implementation because it is anticipated that good 1-way progression on Avenues C and D would be more beneficial to the other arterials in the total network.

Travel Times

Figure 6 shows existing and anticipated vehicle speeds for several key roads in the network. Avenue A should maintain simultaneous signal operation but have the cycle length reduced from 120 to 90 sec. Average speeds should remain the same or be reduced slightly. Avenues C and D, to be changed from 2-way to 1-way operation, should have significant increases in average speed.

Average speeds, after implementation of the proposed changes, will not improve on some of the cross streets. However, improvements should result on the more heavily traveled roads such as Fifth Street. This is achieved at the expense of some other lesser traveled side streets that will likely have speeds for the proposed network slower that those that now exist.

CONCLUSION

The principal deficiency in the SIGOP program, as it related to the specific network analyzed, is that capacity constraint resulting from traffic opposing left turns and conflecting opposing traffic is not accounted for in the computer analysis. This factor has to be estimated manually and, in this case, resulted in a stop and a delay cost adjustment from the computer printout value of $1,460 to $1,780. Furthermore, offsets of the optimized network will likely be somewhat different if this capacity restraint is incorporated into the program as part of the input data.

Coding the SIGOP input data requires considerable work, the amount of time being proportional to the number of signalized intersections coded. To eliminate coding errors and specific problems that develop with each network analyzed often requires several computer runs. This can be overcome with more application of the program, however.

Results obtained from the SIGOP program are valuable to enable the traffic engineer to better understand signal network operations. It saves many man-hours of work not feasible without a computer application of this type. As is the case with most computer work, however, human judgment must be applied in all interpretations of the printouts.

REFERENCES

1. SIGOP: Traffic Signal Optimization Program. Traffic Research Corporation, New York, 1966.
2. SIGOP Traffic Signal Optimization Program Users Manual. Peat, Marwick, Livingston and Company, New York, 1968.

SIMULATION OF CORRIDOR TRAFFIC: THE SCOT MODEL

Edward B. Lieberman, KLD Associates, Inc., Huntington, New York

The increasing activity in controlling access to freeways for the purpose of improving traffic flow has focused on the need to develop control policies for treating the entire corridor network system. This system comprises freeway, servicing ramps, frontage road, and parallel and feeder arterials. It has been observed that ramp metering, while improving conditions on the freeway, can precipitate congestion on the grade roadways. The SCOT model was developed as an evaluative and design tool to predict the performance of alternative control policies and freeway configurations prior to field implementation. A dynamic representation of traffic flow is produced by the model. This paper describes the capabilities and prominent features of the model and some of its representative results.

•IN RECENT years, considerable attention has been focused on the problem of improving the performance of freeway traffic by the application of appropriate ramp controls. During the initial stages of development, such controls have, for the most part, been applied with a view toward improving flow conditions locally, i.e., applying local independent controls of contiguous freeway sections. Congestion on a freeway cannot be relieved (or precluded) unless alternate routes are available that exhibit excess capacity and satisfy the prevailing origin-destination demands. Hence, the corridor control problem involves not only controlling traffic but also routing (or diverting) traffic as well. Local freeway control schemes, which do not consider the system-wide impact of ramp metering (i.e., the response and effect of those "excess" vehicles denied entry to the freeway), may provide limited benefits. Furthermore, freeway congestion, when it does occur, cannot be adequately treated by restricting entry only on the nearby entry ramp.

That a system-wide, network analysis is needed of the entire arterial street-ramp-freeway system has long been recognized (11); such an analysis would lead to the development of an integrated, dynamic control policy for the entire corridor. An attendant need is for a way to evaluate such alternative control policies to determine the most promising candidate prior to the commitment of the substantial resources necessary to implement the control system. Such a tool could also be applied during the planning stage so that future requirements can be determined in the form of additional roadway facilities to satisfy increasing demand levels. It was in response to these needs that the SCOT model was developed.

DESCRIPTION OF SCOT MODEL

The SCOT model represents an evolutionary development based on 2 previous efforts: the UTCS-1 model that simulates urban traffic flow (5) and the DAFT pilot model that simulates corridor traffic (6). The SCOT model represents a synthesis of these 2 models.

The UTCS-1 model is a microscopic simulation of urban traffic, wherein each vehicle is treated as an individual entity as it traverses its path through a network of urban streets and responds to the signal control system, to the presence of other vehicles, and to other well-defined constraints. For each entry link at the periphery of the network, traffic volumes are specified; the routing of traffic is achieved by the specification of vehicle turning movements for each link at each intersection (network node). The trajectory of each vehicle is determined stochastically in response to specified signal control and to the dynamics of the lead vehicle by application of a simplified car-following model. A wide variety of statistical data, expressed in terms of widely ac-

cepted measures of effectiveness, is generated by the model. These data permit the evaluation of alternative signal policies, routing, channelization, parking restrictions, and other control strategies.

The DAFT model is a macroscopic simulation of traffic along a network of freeways, ramps, and arterials. Vehicles are grouped into platoons and lose their individual identities. The platoons are moved along the freeway according to a single, prespecified speed-density relation that applies to all freeway links. Along the nonfreeway links, they travel at the specified free-flow speed for each link and are delayed at the downstream end for a time related to the ratio of green time to signal cycle time (G/C) of the facing signal there and to the volume of traffic on that link. For each entry link at the periphery of the network considered, traffic volumes are specified according to destination node; i.e., the input data consist of origin-destination (O-D) demands that vary with time. The model distributes the resulting platoons of vehicles over the network so as to satisfy these specified O-D patterns and to follow minimum-cost paths. These minimum-cost paths are calculated frequently by the model, based on current conditions. Whenever a platoon reaches a network node, its turning movement there is dictated by its minimum-cost path as it exists at that instant of time. Hence, the model produces a dynamic assignment of traffic as a by-product of the simulation.

The basic approach adopted in the design of the SCOT model was to combine the DAFT and UTCS-1 models appropriately. The microscopic logic of UTCS-1 is applied to those components of the network characterized by signalized, grade intersections. Here, the traffic mechanisms are so complex, because of the many conflicts common to the stop-and-go patterns of urban traffic, that each vehicle must be treated individually and a very small time step must be applied in order to obtain an acceptable level of accuracy in the replication of global traffic flow. Traffic along the freeway, however, has been modeled successfully by the use of fluid-flow analogies and has often been referred to as "stream flow." To apply a detailed, microscopic approach to freeway traffic appears to yield little in the way of additional global accuracy. Furthermore, a microscopic treatment of freeway traffic would greatly magnify computing costs and storage requirements. Hence, freeway traffic is modeled macroscopically in essentially the same manner as the DAFT model.

The SCOT model, then, does "tailor" the degree of detail to the type of network link being processed. Traffic flow on the freeway is described macroscopically, which permits the grouping of vehicles into plattons there and the use of a coarser time step. Traffic on nonfreeway links is treated as a collection of individual vehicles, each processed every second of simulated time. The interface problems associated with this dichotomous logic are resolved as part of the design.

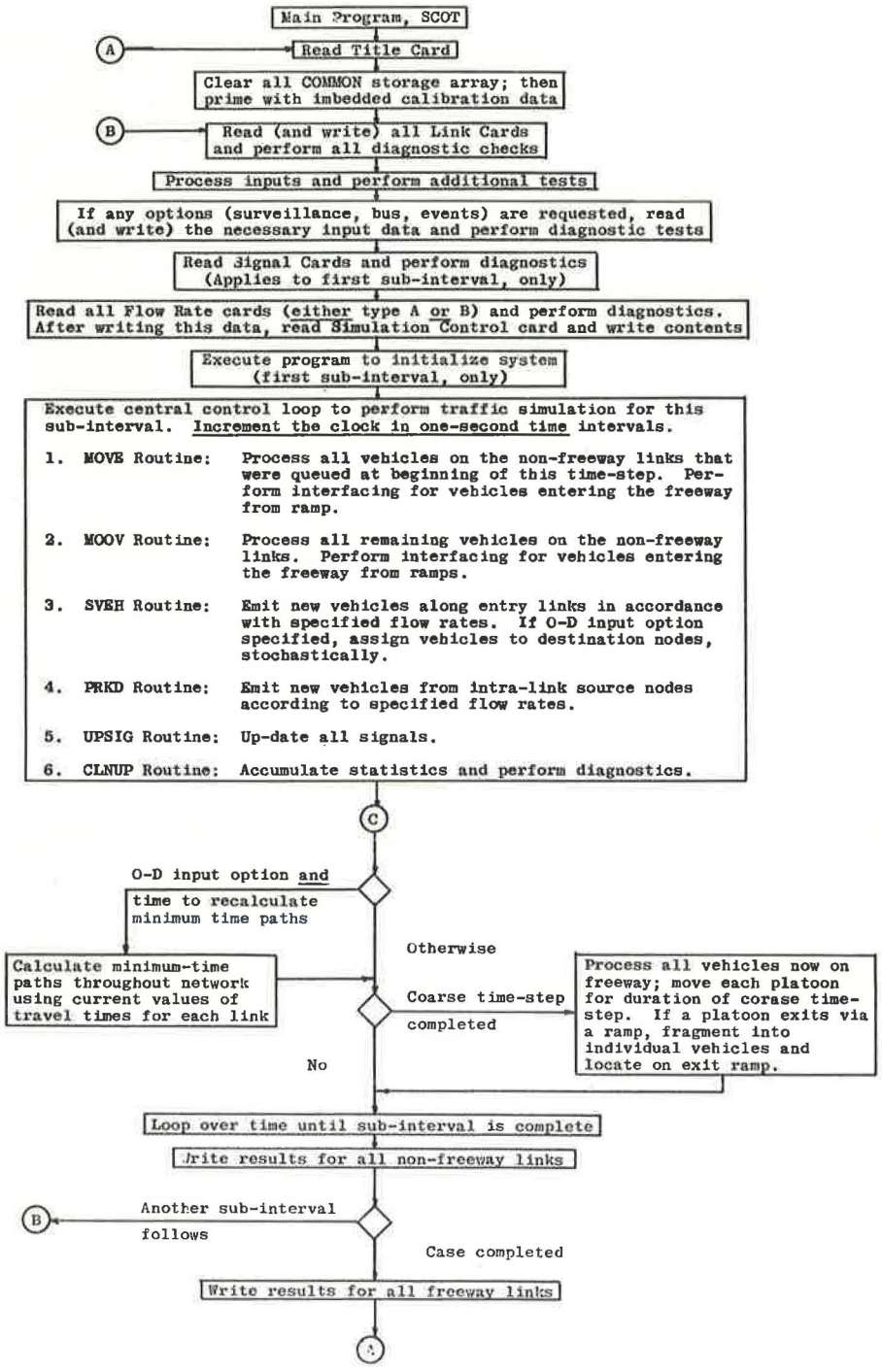
The manner in which traffic demand is specified remained to be resolved; two opposite approaches were applied in the precedent models. The decision was made to incorporate into the SCOT model both approaches in the form of mutually exclusive options open to the user. Hence, either turning movements at each node or O-D volumes may be specified. A flow diagram depicting the global logic of the SCOT model is shown in Figure 1.

Ramp-Freeway Interface

Vehicles traveling along entry ramps are treated as individual entities; when they enter a freeway link, they are treated macroscopically. No problems arise if the logic were content to create 1-vehicle platoons. To do so, however, would compromise the overriding consideration that led to the concept of stream-like platoon flow: conservation of computer storage and execution time. Hence, the SCOT model will attach those vehicles to existing platoons already on the freeway if one is available and is an admissible candidate for such merging.

Availability—An admissible platoon must be located on the freeway at a position that is compatible with the trajectory of the vehicle discharging onto the freeway from an entry ramp. Because the freeway traffic is always processed for a period equal to the coarse time step, the 2 types (freeway and nonfreeway) are usually "out of phase" on the time-scale by an amount ranging from zero to the duration of the coarse time step.

Figure 1. Global logic of SCOT model.



Hence, at the instant a vehicle enters a freeway link from an entry ramp, it must travel along this freeway link at the current link speed for a time that would erase this phase difference. The logic then examines those platoons located on the freeway link to locate one nearby that is admissible for the vehicle to join. If one is found, the vehicle joins the platoon, retaining its identity in a logical sense. The statistics of the platoon are adjusted to reflect its new member. The result of this logic is to reduce by one the number of individual entities (vehicles and platoons) that must be stored and processed by the model. If an admissible platoon is not located, there is no recourse but to create a 1-vehicle platoon.

Admissibility—The definition of this term depends on the input option that is active. If turning movements at each node are specified, then any nearby platoon is admissible. If O-D volumes are input, then only those platoons having the same destination node are candidates for accepting the vehicle. (Another factor related to efficient use of computer storage also impacts on admissibility.)

Freeway-Ramp Interface

When a platoon of vehicles on the freeway is directed to an exit ramp, it must be fragmented into its component parts (individual vehicles). Furthermore, the phase difference noted above must be taken into account. This task is relatively straightforward; the speeds of these vehicles are adjusted so that any effects due to the difference between the coarse and fine time steps are rapidly resolved. The net effect is to preserve the integrity of the vehicle trajectories and the attendant statistics.

Freeway-Freeway Interface

A platoon of vehicles traveling along the freeway from link to link will experience varying speeds depending on whether it is within a link or at the junction of 2 freeway links.

Consider a platoon that is located in the midst of a link at the conclusion of the coarse time step. Before that platoon "resumes" motion, a calculation to ascertain the current mean speed has been completed. Hence, this platoon proceeds to move along the same link but at a speed reflecting traffic conditions at the beginning of this new coarse time step.

A platoon that arrives at the downstream terminus of a freeway link in the midst of the current time step abruptly assumes the mean speed associated with the receiving link for the remainder of this coarse time step. If traffic on the receiving link is congested to the extent that jam density is realized, the platoon is halted and does not enter the next link.

Minimum-Time Paths

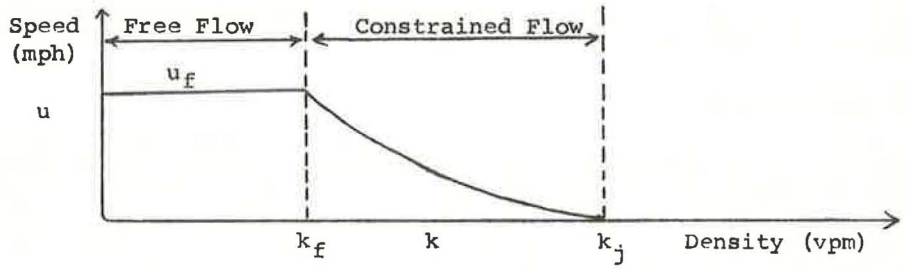
When the input option that specified O-D volumes on the entry links is active, the routing of traffic through the network reflects dynamically changing minimum-time paths. The time interval between successive calculation of these paths is specified as input. Because these calculations require the spooling of data to and from disk storage and because the minimum-time paths are not particularly sensitive to short-term changes in traffic performance, this interval is an integer multiple of the coarse time step used for moving freeway traffic. The algorithm adapted for the calculation of these minimum travel-time paths is given in another report (7).

Speed-Density Formulation

The speed-density formulation, which forms the crux of the freeway component of the SCOT model, is patterned after the work of May and Keller (1). Their work reviewed the existing integer traffic models that had been derived from the general car-following equation given earlier (2). They then presented a methodology for determining a more general, noninteger solution to the car-following equation. In a later paper (3) they extended the approach to 2-regime representations.

As indicated in the later paper (3), the "best" solutions applied to a sample of freeway data exhibited a rather poor fit in the region of a density of fewer than 60 vehicles/mile. These relatively poor results should not be surprising. It is well recognized that the car-following law does not apply during periods of light flow, reflecting the lack of interaction among the widely separated vehicles. It has been observed that vehicles spaced 200 ft apart or more behave essentially independently of one another. Hence, the car-following model, in general, would not apply for densities of fewer than 25 vehicles/mile.

The adoption of the following 2-regime model provides greater accuracy for the resulting curve-fit function relating speed and density:



As shown, the car-following model will be applied in the constrained flow regime; a mean free-flow speed will apply when traffic is sparse. This approach is consistent with other data (4). The car-following equation utilized is

$$\ddot{x}_{n+1}(t+T) = \alpha \frac{\dot{x}_{n+1}^{m}(t+T)}{[x_n(t) - x_{n+1}(t)]^{\ell}} [\dot{x}_n(t) - \dot{x}_{n+1}(t)] \qquad (1)$$

where

x_n = position of lead vehicle, n;
x_{n+1} = position of following vehicle, n + 1;
T = response time lag of following vehicle;
m, ℓ = exponent parameter;
α = coefficient;
t = independent variable, time; and
$\dot{x}$, $\ddot{x}$ = speed and acceleration respectively.

On the basis of previous work (1), we will assert that $m < 1$ and $\ell > 0$ but $\neq 1$. For this admissible range of m and ℓ, the solution to Eq. 1 takes the form

$$u = u_f\{1 - [(k^a - K_f)/(K_j - K_f)]\}^b \qquad (2)$$

where

$a = \ell - 1$;
$b = (1 - m)^{-1}$;
k_f = density demarking juncture of free-flow and constrained-flow regimes, vehicles/lane-mile;
k_j = jam density;
u_f = mean free-flow speed, mph;
$K_f = (k_f)^a$;
$K_j = (k_j)^a$;
k = current value of density in the range, $k_f \leq k \leq k_j$, where, in general, $k_f > 0$, and if $k < k_f$, then $u = u_f$.

The optimal values of density and speed may be written (8) as

$$k_{opt} = k_j[1/(1 + ab)]^{1/a} \quad (3)$$

$$u_{opt} = D\, u_f[ab/(ab + 1)]^b \quad (4)$$

where $D = [1 - (K_f/K_j)]^{-b}$.

The formulation given above is valid as long as $k_{opt} > k_f$; i.e., maximum flow occurs within the constrained-flow regime. This condition is satisfied if

$$a \cdot b < (K_j - K_f)/K_f \quad (5)$$

The maximum flow rate, Q_{max}, is the product of Eqs. 3 and 4.

Calibration of Speed-Density Relation—As shown in Figure 2, the corridor network comprises unidirectional links and nodes. These nodes may represent either signalized intersections or juncture points of contiguous freeway links and ramps. The partitioning of the freeway into a linear network is related to changes in topology from one link to the next: change in grade, number of lanes, horizontal curvature, indeed, any factor that would produce a change in the speed-density characteristics of traffic. This feature provides the model with the capability of predicting the performance of traffic over a freeway characterized by rolling terrain, variation in width (due to, e.g., a lane drop or a stopped vehicle), or any factor. The impact of inclement weather may be determined as well.

It is necessary, then, to calibrate this speed-density relation for each category of freeway links. If 2 (noncontiguous) freeway links exhibit the same global geometric and traffic characteristics (grade, length, width, and traffic composition), it is reasonable to assume that both links may be described by the same relation. Hence, links displaying similar characteristics may be grouped within a single category. There are a total of 5 parameters required for calibration: u_f, k_f, k_j, a, and b. An analysis has been developed, and a computer program has been written that accepts experimental data in the form of speed-density measurements for each link and yields that set of 5 parameters that produces the best least squares fit consistent with the measured capacity, Q_{max}, of the freeway link (9).

When a sufficiently large inventory of such calibrations has been accumulated, it may no longer be necessary to conduct experiments to acquire speed-density data. It could then prove feasible to study projected freeway designs, utilizing this inventory of calibration for the purpose of evaluating each candidate design. In this mode of usage, the model takes on the attributes of a design tool that can help to optimize, on an overall cost-effective basis, the design of future freeway configurations.

APPLICATION OF MODEL

A series of sensitivity tests was performed to illustrate the application of the model while the calibration and validation effort was being completed. Figure 2 shows a network that represents a portion of the Central Expressway north of Dallas. Traffic is introduced into the network along the peripheral (entry) links and is routed through the network according to specified turning movements at each node.

Three conditions were studied. The demand at the upstream (southerly) node of the freeway was varied; volumes of 3,400, 3,800, and 4,200 vph were introduced. The individual freeway links downstream experience varied demand levels because of the presence of both entry and exit ramps. The results of these exercises are given in Table 1 and shown in Figure 3.

The first increment of 400 vph has little impact on most link travel times; demand does not exceed link capacity on any link. When volume increases to 4,200 vph at the upstream node, several links experience congested conditions resulting in vehicles queuing on the freeway. The response of traffic on the freeway is nonuniform; i.e.,

Figure 2. Elements of freeway corridor.

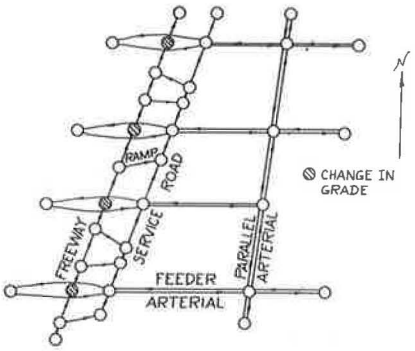


Figure 3. Travel time versus distance for various demand volumes.

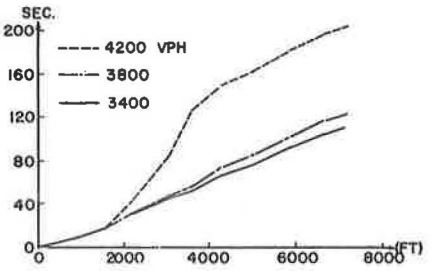


Table 1. Link travel times (sec) based on demand (vph) at upstream node.

		3,800 vph		4,200 vph	
Link	3,400 vph	Number	Percent[a]	Number	Percent[a]
1	9.4	9.4	0	9.4	0
2	8.3	8.3	0	8.3	0
3	12.5	13.3	5	25.3	102
4	14.4	15.5	8	44.9	211
5	7.1	8.9	25	38.5	443
6	14.4	18.6	13	23.9	66
7	10.4	11.1	7	11.1	7
8	16.9	16.9	0	21.5	27
9	12.8	13.4	5	13.3	4
10	6.9	7.1	3	7.1	3

[a]Change from the 3,400-vph demand level.

Figure 4. Traffic volume response to freeway blockage.

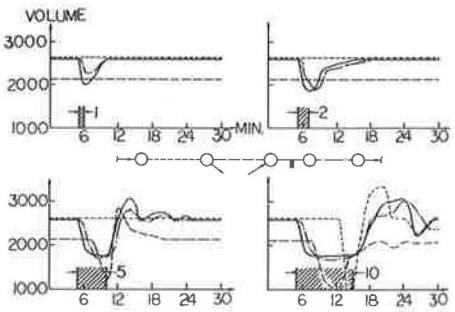


Figure 5. Traffic speed response to freeway blockage.

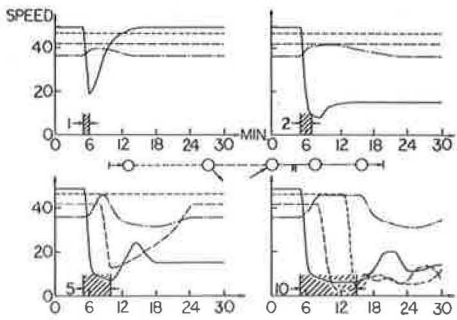
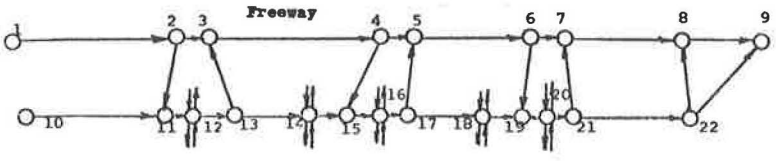


Figure 6. Network configuration.



not all links exhibit the same behavioral characteristics. It is this ability of the model to identify those links that impact most heavily on system performance that is of considerable value to the engineer.

Another series of exercises was performed to test the ability of the model to respond properly to abrupt changes in conditions. Although the scenarios presented below do not fall within the intended scope of the model (in the sense that freeway traffic is treated macroscopically), it is nevertheless assuring that the model remains stable and that the results appear to be reasonable.

Four exercises were conducted on a small network comprising 4 freeway links and 2 ramps. A transient blockage, representing an "incident," was introduced on one freeway link as shown in Figure 4 (solid line). Prior to the blockage, the freeway volume was well below capacity. This blockage, which removes one of the 2 available lanes from service, prevails for 1, 2, 5, and 10 min in duration. The response of traffic for each case is shown in Figures 4 and 5.

A blockage of short duration produces a primarily local impact on traffic performance: Traffic performance on the afflicted link deteriorates rapidly, while the downstream (receiving) link experiences a "relief" that is reflected in decreased travel times. As the blockage persists the resulting shock wave propagates upstream and causes queuing on the feeder links as well. A larger portion of the network is affected; not only the freeway but also the ramp and frontage road operations (not displayed) reflect the breakdown caused by this blockage. Recovery is slow, congested conditions prevail long after the cause is removed, and oscillatory flow conditions characterize the system. When the demand persists at the same high level (but below capacity), forced-flow conditions (reduced speed) may be retained on recovery. Only a slackening in demand will return the freeway to free-flow operations.

The test of any corridor control and route diversion system is its ability to cope with conditions that occur abruptly and spread rapidly to engulf large portions of a transportation network. The ability of the model to accommodate such a condition represents a valuable asset.

Present plans call for the SCOT model to be implemented to replicate the ramp-metering policy now in effect on the North Central Expressway in Dallas, but these results are not as yet available. This application of the model is illustrated by results obtained with the precedent DAFT model (6). The network configuration shown in Figure 6 was utilized for this purpose. Incoming traffic was specified in terms of O-D volumes that varied with time. The following data for traffic demand leaving node 1 and destined for node 9 illustrate a typical entry link specification:

Time	Volume (vph)
7:30	1,770
7:40	2,000
7:52	2,320
8:04	2,320
8:16	2,000
8:28	1,770

Nodes 1, 10, 12, 16, and 20 act as origins (sources) only; node 9 acts as a destination (sink) only; and nodes 14 and 18 act as both origins and destinations. Signals control traffic at nodes 12, 14, 16, 18, and 20.

Mean free-flow speed along the freeway (nodes 1 through 9) is specified as 55 mph; along the service road (nodes 10 through 22), free-flow speed is 45 mph for the longer links and 40 mph for the shorter ones; along the freeway ramps, a value of 40 mph is applied.

Specification of Alternative Policies

Four cases were analyzed:

1. Freeway has 2 lanes from node 1 to node 9 and no ramp control;

2. Freeway has 2 lanes and moderate ramp control;
3. Freeway has 2 lanes and strict ramp control (i.e., the G/C for the ramps is less than that for case 2); and
4. Freeway is widened to 3 lanes between nodes 5 and 9 and has no ramp control.

All 4 cases were subject to the same traffic demand. The program was executed, and the results were tabulated for 1 hour for the time from 7:30 to 8:30.

History of Average Speeds Along Freeway

Figure 7 shows average speed as a function of time along the various links constituting the freeway. The pattern of demand volumes is such that the 2 upstream links are not affected by changes in control policy.

The subsequent links (3, 4) and (4, 5) are moderately influenced by control changes; both links experience a decay in speed the first half hour and then a relaxing of congestion as demand tapers off.

For the case of no ramp control, link (5, 6) rapidly becomes a bottleneck, with traffic flow reduced to crawl-speed conditions. The consequent relief realized by the downstream link reflects the choking of traffic flow that creates a condition of excess capacity downstream.

The moderate ramp control in this case merely served to delay the onset of congestion, not to prevent it. The traffic performance for this case is similar to the previous one, except for a shift of about 8 minutes. This behavior confirms the need for a careful design of a ramp-control policy to cope with actual traffic conditions; an inadequate design will yield minimal benefits.

The application of strict ramp controls (case 3) radically alters the character of traffic flow. The inlet ramps become saturated quickly, causing traffic to be directed to the service road prior to the onset of freeway congestion. As is shown, the freeway traffic is only moderately affected by the heavy peak-hour demand, and minimum speed for this case is again shifted somewhat with respect to the previous one.

As expected, the addition of another lane greatly improves matters on the affected links. It is important to note, however, that this increase in capacity acts to draw onto the entire freeway additional demand that could create bottlenecks upstream, as indicated for link (3, 4).

Flow Rates Along Freeway

The flow rates on each freeway link are shown in Figure 8 as ratios referenced to maximum flow rate (capacity) to illustrate the utilization of the facility. Formation of a bottleneck not only chokes flow at the constricted section but also creates an excess of capacity downstream.

Freeway Travel Time

Figure 9 displays the freeway travel time between nodes 1 and 9 for vehicles leaving node 1 at various clock times. A properly designed control system can reduce travel time as much as 50 percent, compared with a system with no control.

Ramp Traffic

Traffic density along the entry ramps (13, 3) and (17, 5) are shown in Figure 10. The impact of ramp control is clearly illustrated. Ramp (13, 3) rapidly becomes saturated under strict control, and traffic diverts onto the service road.

Ramp (17, 5) feeds freeway link (5, 6), which experiences bottleneck conditions. As a consequence of this congestion, the minimum-time path from node 17 to node 9 lies along the service road—not the freeway. Hence, the sharp drop in ramp occupancy for cases 1 and 2 reflects the onset of freeway congestion and the resulting diversion of traffic to the service road. The addition of a third lane on the freeway (case 4) is reflected in increased usage of this entry ramp.

Figure 7. History of freeway speed.

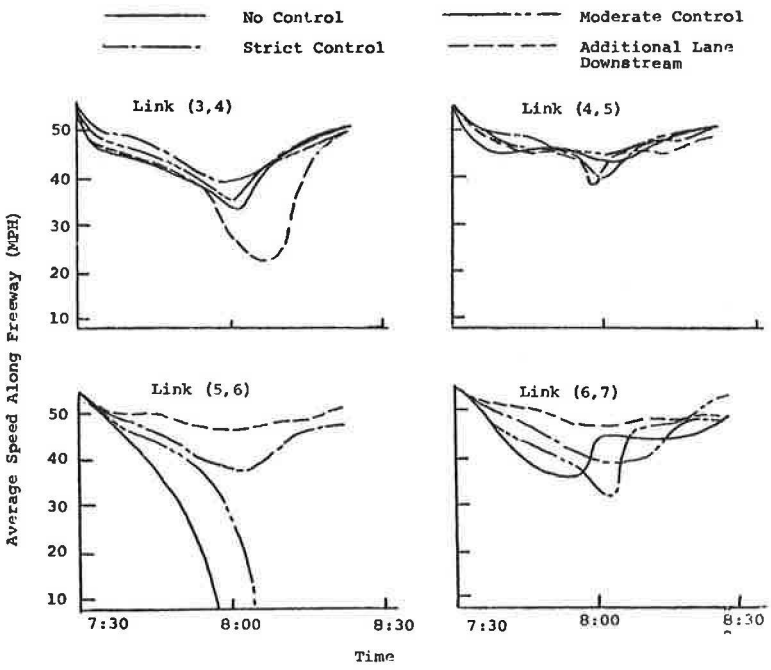


Figure 8. History of flow rates along freeway.

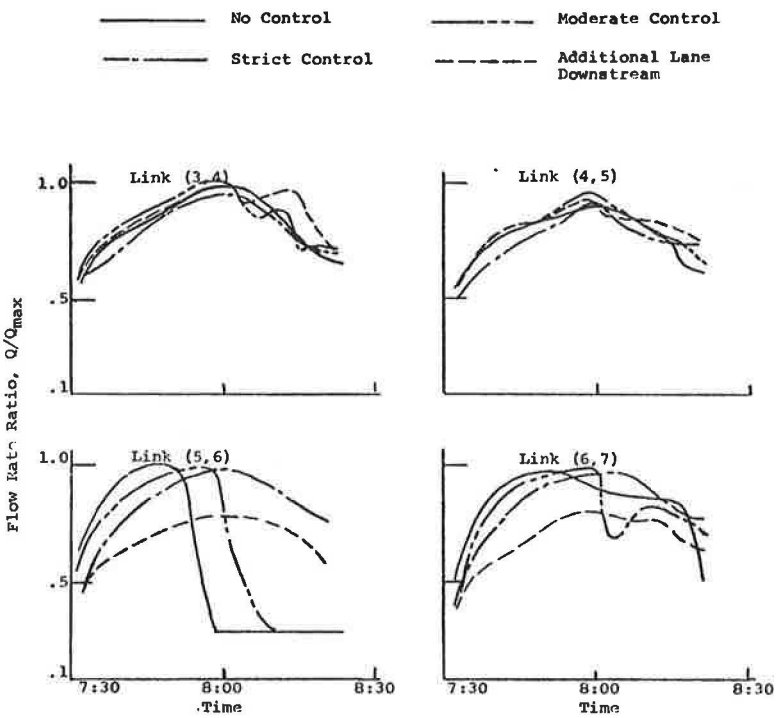


Figure 9. Travel time between nodes 1 and 9.

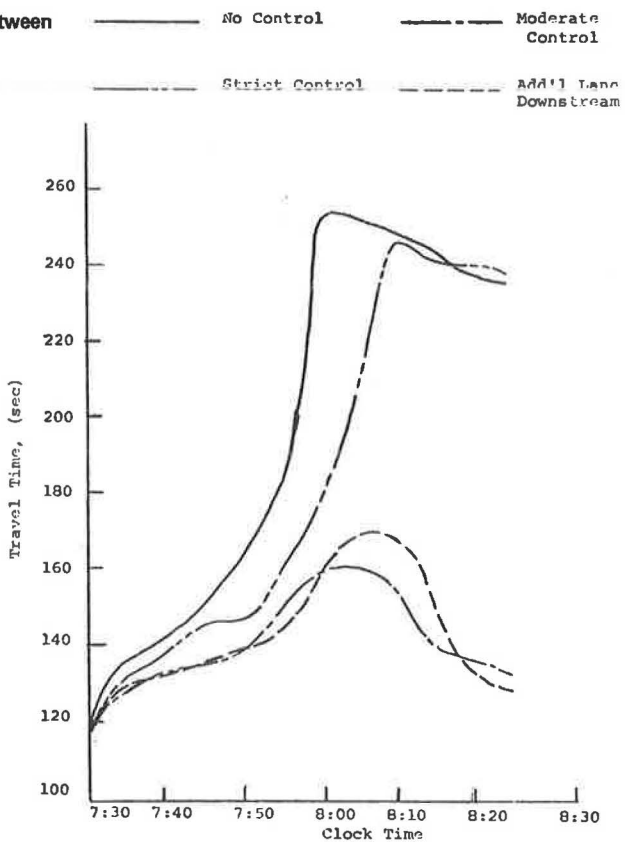
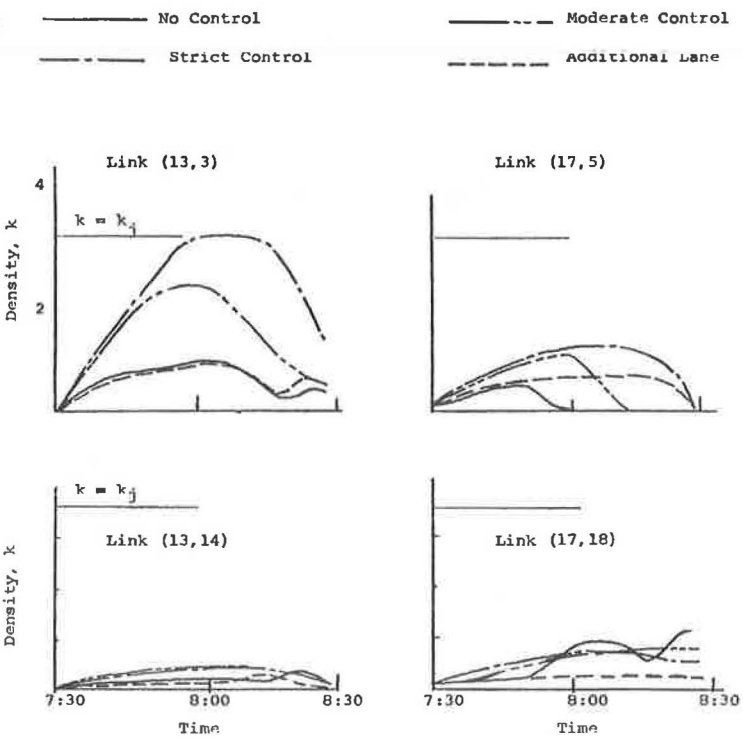


Figure 10. Traffic on ramps and on service road.



Service-Road Traffic

As expected, the presence of ramp control increases the density of traffic along the service road. Interestingly, this increase does not unduly tax the capacity of the service road, confirming observations elsewhere (10). In fact, as shown for link (17, 18), maximum density resulted from the diversion of traffic because of congestion and not ramp control. The service road is severely underutilized when a third lane is added.

SUMMARY

It has been stated (12), "...there is growing evidence that research in metering and controlling ramp traffic must encompass a comprehensive approach to the total capacity-demand of the freeway and the existing street network treated as a single system." This is particularly true for those corridor systems that are considered for computerized control. The prospect of improving freeway operations at the expense of creating long queues on entry ramps (10) and of disrupting traffic along neighboring arterials is to be avoided at all cost.

The SCOT model was developed as a vehicle for testing real-time control policies for an entire corridor: freeway ramps, frontage roads, and adjoining feeder and parallel arterials. The simulation approach makes it possible to study in detail the dynamic traffic responsiveness to an on-line, computer-controlled system. Those control policies that produce lesser benefits relative to others may be discarded early in the design stage. Incremental cost-benefit analyses may be conducted to determine at what stage an increase in system sophistication and cost is no longer justified by a commensurate improvement in traffic performance. Similar analyses may be performed in the design stage to assist in the determination of the vertical profile of the freeway and of the placement of freeway ramps, signal controls, and detectors.

REFERENCES

1. May, A., Jr., and Keller, H. Non-Integer Car-Following Models. Highway Research Record 199, 1967, pp. 19-32.
2. Gazis, D., Herman, R., and Rothery, R. Non-Linear Follow-the-Leader Models of Traffic Flow. Operations Research Jour., Vol. 9, 1961.
3. May, A., Jr., and Keller, H. Evaluation of Single- and Multi-Regime Traffic Flow Models. Fourth Internat. Symposium on Theory of Traffic Flow, Karlsruhe, June 1968.
4. Courage, K. Some Electronic Measurements of Macroscopic Traffic Characteristics on a Multilane Freeway. Highway Research Record 279, 1969, pp. 107-120.
5. Bruggeman, J., Lieberman, E., and Worrall, R. Network Flow Simulation for Urban Traffic Control System. KLD Assoc., Inc., Tech. Rept. FH-11-7462-2, 1971.
6. Lieberman, E. Dynamic Analysis of Freeway Corridor Traffic. Joint Eng. Conf. in Chicago, Paper 70-TRAN-42, Oct. 1970.
7. Hu, T. C. Revised Matrix Algorithms for Shortest Paths in a Network. Jour. of SIAM, Vol. 15, No. 1, 1967.
8. Lieberman, E. Logical Design of the SCOT Model. KLD Assoc., Inc., Tech. Rept. 2, 1971.
9. Lieberman, E. Calibration of Freeway Speed-Density Relation. KLD Assoc., Inc., Tech. Memo 1, 1971.
10. Courage, K. Evaluation and Improvement of Freeway Corridor Operations. Paper presented at Highway Research Board 49th Annual Meeting, Jan. 1970.
11. Wattleworth, J. Peak-Period Analysis and Control of a Freeway System. Highway Research Record 157, 1967, pp. 1-21.
12. Cribbins, P. D. Discussion of paper by Drew, D. R., Gap Acceptance Characteristics for Ramp-Freeway Surveillance and Control. Highway Research Record 157, 1967, pp. 135-136.

LOGICAL DESIGN AND DEMONSTRATION OF UTCS-1 NETWORK SIMULATION MODEL

Edward B. Lieberman, KLD Associates, Inc., Huntington, New York; and
Richard D. Worrall and J. M. Bruggeman, Peat, Marwick, Mitchell and Company, Washington, D. C.

A description is given of a microscopic simulation model designed as an evaluative tool for urban traffic control policies. The need for such a tool is explored, and the underlying benefits of the simulation approach are discussed. The logical structure of the model is described; the input requirements and statistical output generated by this FORTRAN-coded program are detailed, and samples are illustrated. The prominent features of the model are examined in detail.

•CONSIDERABLE attention has been devoted in recent years to the development of sophisticated control policies for application to networks of urban traffic signals. The need for such advanced control policies is evidenced by the growing volume of vehicular and pedestrian traffic in urban environs and the resulting congested conditions that threaten the viability of urban transportation systems. Improving the effectiveness of traffic control appears to hold the greatest near-term potential for relieving those conditions that impede the flow of traffic and act to constrict the economic and social activities of urban areas.

The design of such control policies and their implementation on urban networks represent a considerable investment in manpower and in hardware. The implemented control policy, the traffic demand, and the urban roadways constitute a complex system that is not amenable to closed-form mathematical analysis. The design of the various components of this system—surveillance detectors, communication network, signal controllers, digital computer configuration, control policy design, software package—interact with one another in complex ways that are difficult to evaluate a priori.

It is not practicable to adopt a "try-something-and-see" policy for such a system. The selection of a digital computer and the choice of detectors (their number and placement) are decisions involving a considerable capital investment, which are strongly related to the design of the control policy to be implemented. It appears, then, that some analytical technique is required that can evaluate alternative control policy designs for the environment under consideration. The development of the UTCS-1 simulation model is an outgrowth of the need to fill this requirement.

The design of the UTCS-1 model for the study of urban traffic flow and dynamic signal control systems was directed to satisfy the following objectives: The model must accurately describe the real-world dynamics of urban traffic and respond to a wide variety of controls, including responsive systems actuated by an on-line digital computer. This description of traffic dynamics must be expressed in terms of significant traffic parameters and measures of effectiveness that characterize the performance of each component (link and node) of the network. The accuracy and reliability of these results must satisfy the basic research objective of utilizing the model as a diagnostic engineering tool for the evaluation of alternative policies of traffic control, channelizations of traffic, and turning and parking restrictions for any urban network configuration and composition of traffic, including buses.

The model must accurately describe individual responses to the constraints imposed by the control system, geometrical configuration, competing traffic elements (vehicular and pedestrian), and operating characteristics of the vehicle. The vehicles within the system must be treated as a collection of independently motivated entities, interacting

with the urban network environment. Hence, a discrete, stochastic, microscopic simulation model is required.

Such a model requires a balance between considerations of sufficient detail to accurately describe the system and considerations of practical utility, e.g., cost of operation and need to satisfy storage restrictions of available computers. Furthermore, to promote wide acceptance of the resulting computer program requires that a high-level programming language acceptable to all computers regardless of manufacturer be employed. To further promote the use of the program requires that the substantial costs of data acquisition and preparation be reduced to a minimum. In addition, the program must be structured as a logical synthesis of closed routines such that most modifications will affect a minimum number of these routines.

The UTCS-1 program was designed to satisfy the requirements and objectives described above and was oriented to serve the needs of the practicing traffic engineer. The model also serves the needs of future research activities, the design of surveillance systems, and the development of computer software to implement traffic control policies.

The following subsections include a discussion of computer simulation and a detailed exposition of the features that are embodied in the UTCS-1 model (1).

RATIONALE FOR COMPUTER SIMULATION

Simulation is a modeling process that permits the study of a complex, dynamic system that cannot be adequately described by mathermatical relations alone. It is of particular value when the system under study comprises components that are highly variable and must be described in probabilistic terms through the medium of distribution functions. The advantage of a simulation model, properly calibrated and validated, is its use as a test-bed to economically study the system's dynamic response to a variety of input specifications.

The general attributes of simulation models may be summarized as follows:

1. Simulation permits the study and experimentation of the complex internal interactions of a system;
2. Detailed study of the simulation results provides a better understanding of the system and leads to improvements in its operation;
3. Simulation of a complex system yields valuable insight into those components that most profoundly affect the system's performance, identifies those that are less sensitive in this respect, and promotes an understanding of how these components interact with one another;
4. A simulation model may be used to experiment with new concepts that have yet to be put into practice and to test new policies and decision rules for operating a system before resources are committed for developing them;
5. The process of creating the simulation model requires that the system be "fragmented" into its component parts, each of which may be amenable to successful analytic formulations;
6. A simulation study is less costly than most other forms of experimentation; and
7. Simulation is safe, for the system can be tested by the computer rather than in a real situation where the possible creation of hazardous conditions would be infeasible.

A simulation model may be classified as either microscopic or macroscopic. The latter type is designed to yield the global behavior patterns of the system under study and does not require the detailed mechanisms of each component to be deeply probed as does the former type. In other cases, the intrinsic assumptions associated with the design of a macroscopic model completely invalidate its integrity. For an urban traffic system, it is mandatory that the model be designed to properly replicate the microscopic dynamics of the system component, in order to produce results that accurately reflect real-world behavior.

In summary, a system that functions dynamically as a complex logical interaction of many small components, some of which are stochastic and are characterized by rapidly changing conditions, is a prime candidate for simulation modeling for the purpose of analysis and eventual design.

FEATURES OF UTCS-1 MODEL

The UTCS-1 model has the following features:

1. Representation of any urban network having as many as 100 intersections and associated input and output links;
2. Simulation of the full range of geometric configurations found in the central areas of U.S. cities of 50,000 population or more;
3. Replication of traffic performance under all forms of network traffic control, including dynamic control systems based on the area-wide surveillance of traffic parameters such as traffic volume, average speed, traffic density-vehicle content, and smoothness of flow;
4. Provision for a flexible mix of both input and output options (standard output measures that may be accumulated by cycle or any larger time period for both an individual intersection and the network as a whole include link volume, average link speed, average link travel time, total vehicle-miles of travel, traffic density-vehicle content, vehicle delay, queue length, number of stops, and smoothness of flow);
5. Stochastic simulation of individual vehicles by type, utilizing a simplified car-following model, time-scanning methods, and 1-sec memorandum notation;
6. Detailed treatment of both intersection and intralink behavior, including queue discharge and turning behavior, intersection "spillback," pedestrian-vehicular traffic interaction, and intralink acceleration and deceleration;
7. Treatment of nonsignalized controls, e.g., STOP and YIELD signs, free-force merge, and exclusive channelization (e.g., turning lanes);
8. Treatment of bus traffic, including specification of bus routes, bus stop-dwell times, and bus-vehicular traffic interaction;
9. Treatment of intralink friction, including both legal and illegal parking, parking garage flows, and intralink rare events (e.g., cab pickups, delivery vehicles, and temporary lane closure);
10. Simulation of alternative surveillance systems, including location and type of detector (e.g., presence detector, counter, and spot speed detector), detector signals, and detector reliability (as many as 3 detectors may be utilized per lane and as many as 12 per link);
11. Provision for minimum essential input data to simulate a given set of field conditions, including allowance for both location-specific and phenomenological (i.e., network-wide) data inputs, and provision for a comprehensive set of default options-values;
12. Description of traffic network at varying levels of detail and specificity; and
13. Provision for internal control of network "fill-time."

MODEL STRUCTURE

An urban street network is decomposed for the purpose of model operation into a network of unidirectional links (streets) and nodes (intersections), as shown in Figure 1. Each link may contain as many as 5 moving lanes. Wider streets may be accommodated by an individual street being separated into 2 successive links. Intersection control may take the form of 9 different signal configurations, including STOP and YIELD sign control. As many as 9 signal phases may be incorporated into any given signal cycle. Provision is made for the simulation of alternative surveillance systems, incorporating as many as 3 detectors in any 1 lane and as many as 12 detectors in total for any 1 link. The information derived from these detectors may in turn be used as input to alternative traffic-responsive network control strategies.

Vehicles are emitted onto the network along entry links or from "source" nodes located within the interior of the network. In addition, "sink" nodes may extract traffic from specified links within the network. These source and sink nodes serve to represent the operation of parking lots, garages, or service stations or of intermediate streets not represented on the full simulated network. Traffic is discharged from the network via exit links or through the sink nodes described above.

Each vehicle is uniquely identified in terms of its performance characteristics and is simulated as an individual entity. Its time-space trajectory as it traverses the network is recorded to a resolution of 0.1 sec. Associated with each vehicle is a table that is updated every second, so that a record is maintained of the vehicle's statistical history. This record includes specification of the type of vehicle (e.g., automobile, truck, or bus), cumulative trip time, cumulative distance traveled, cumulative delay incurred, current lane and link position, projected turning movement at the downstream intersection, number of stops incurred to date, and current position in queue. Hence, the complete trajectory history and the current status of each vehicle on the network are known at any given point in time. The data contained within these "vehicle arrays" are interrogated by the various subroutines in the model and are also periodically accumulated and displayed as output.

The simulation of separate vehicles as individual entities permits the accurate formulation of a series of microscopic components of the model dealing with phenomena such as turning logic at intersections, queue discharge, and lane-switching. This in turn is the basis for the provision within the model system of a wide spectrum of features that otherwise could not be treated directly, e.g., the treatment of bus traffic with specified routes and midblock bus stops, and that are essential if the model is to accurately reflect the response of real-world traffic behavior to alternative network control strategies.

The model is programmed completely in FORTRAN IV and is operational on both CDC 6600 and IBM 360 hardware. It consists of 43 separate subroutines, including a central executive routine, UTCS-1, which is used to activate the other subroutines in logical sequence. The central UTCS-1 routine is executed once each "time step" throughout the simulation. Provision is also made for updating selected inputs to the model, including specification of signal-timing and input-flow rates at the start of each of a series of regular "subintervals" within the overall simulation run. Figure 2 shows the logical flow for the main UTCS-1 routine.

INPUT REQUIREMENTS

The model has been designed to meet a number of specific objectives concerning input data requirements. These include the following:

1. Minimization of the necessary information required for the simulation of a given network or a given set of traffic conditions;
2. Utilization, to the maximum possible extent, of data that are normally available in the files of traffic engineering departments;
3. Through the medium of input options, restriction of input data to those that are directly relevant to the needs of the user;
4. Provision for the effective utilization of varying levels of input data, depending on the quantity and the quality of the information available to the user; and
5. Provision for default information for use in those instances where specific calibration information cannot be obtained for a particular network.

Within the framework defined by these objectives, the necessary set of input data required to operate the model may be grouped under 2 separate headings: location-specific parameters, reflecting the particular characteristics of a given network link or intersection, and phenomenological measures, which are assumed to remain constant across the entire network to be simulated, regardless of the location of a specific link or intersection.

Specifically, the following input data are required:

1. For each network link, number of moving lanes, length, capacity of left-turn pocket, desired free-flow speed, mean queue discharge rate, turning movements at downstream node, identification of receiving links, pedestrian volume, and lane channelization;
2. At each intersection complete specification of signal (or sign) control, including sequence and duration of each phase and identification of signal facing each approach;

Figure 1. Validation network.

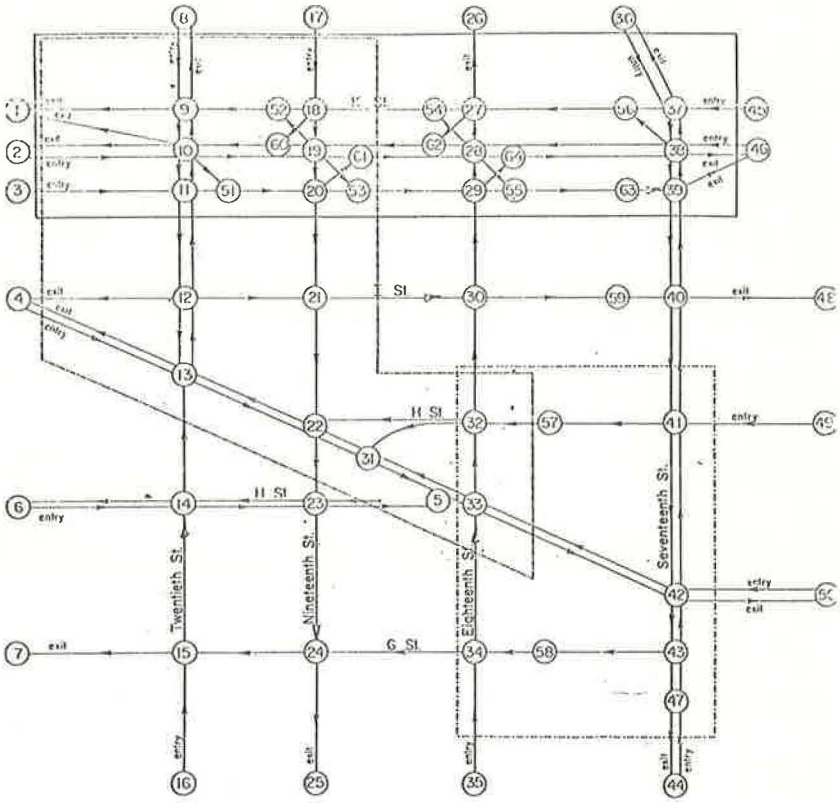


Figure 2. Logic flow for UTCS-1 executive routine.

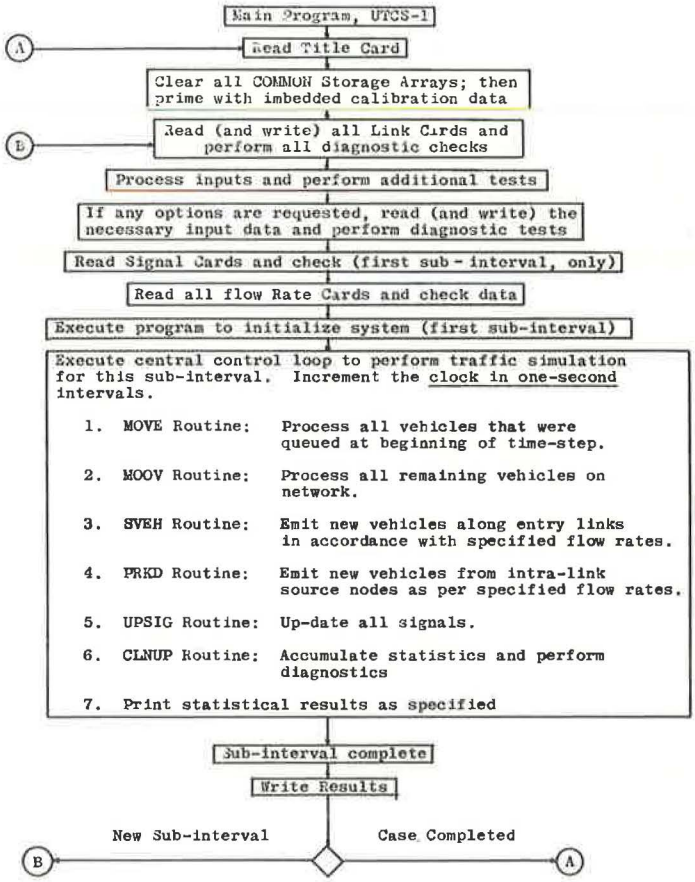


Figure 3. Link input data.

LINK	LANE	SPAN	MEAN LTRN	U	F	H	TURNING INCIDENCES LEFT	THRU	RIGHT	DIAG	DESTINATION NODES LEFT	THRU	RIGHT	DIAG	PED LOST	PED DEN	LANE USAGE 1	2	3	4	5	TYPE	L
(23, 16)	2	86	0		24	24	0	100	0	0	0	27	0	0	0	0	0	0	0	0	0	1	19
(16, 27)	2	590	0		28	24	0	59	41	0	0	14	26	0	0	0	4	0	0	0	0	1	20
(27, 14)	2	50	0		28	24	0	100	0	0	0	18	0	0	0	0	0	0	0	0	0	1	21
(14, 18)	2	447	0		28	24	0	100	0	0	19	6	0	0	0	0	0	0	0	0	0	1	22
(18, 6)	2	50	0		28	24	0	100	0	0	0	9	0	0	0	0	0	0	0	0	0	1	23
(6, 9)	2	353	0		28	24	0	46	54	0	10	1	8	0	0	0	4	0	0	0	0	1	24
(24, 28)	2	645	0		28	24	0	86	0	14	0	19	27	14	0	0	0	0	0	0	0	1	25
(28, 19)	2	497	0		28	24	0	100	0	0	20	10	0	6	0	0	0	0	0	0	0	1	26
(19, 10)	2	403	0		28	24	0	91	0	9	11	2	9	1	0	0	0	0	0	0	0	1	27
(10, 19)	2	415	0		28	24	0	100	0	0	0	28	20	7	0	0	0	0	0	0	0	1	28
(19, 28)	2	495	0		28	24	0	92	0	8	27	24	0	15	0	0	0	0	0	0	0	1	29
(28, 24)	2	620	0		28	24	0	100	0	0	23	32	25	0	0	0	0	0	0	0	0	1	30
(11, 5)	2	50	0		28	24	0	100	0	0	0	20	0	0	0	0	0	0	0	0	0	1	31
(5, 20)	2	365	0		28	24	0	32	68	0	0	7	21	0	0	0	4	0	0	0	0	1	32
(20, 7)	2	50	0		28	24	0	100	0	0	0	29	0	0	0	0	0	0	0	0	0	1	33
(7, 29)	2	445	0		28	24	0	100	0	0	28	15	0	0	0	0	0	0	0	0	0	1	34
(29, 15)	2	50	0		28	24	0	100	0	0	0	34	0	0	0	0	0	0	0	0	0	1	35
(34, 25)	2	250	0		28	24	0	50	50	0	24	32	33	0	0	0	4	0	0	0	0	1	36
(15, 34)	2	320	0		28	24	0	100	0	0	0	25	0	0	0	0	0	0	0	0	0	1	37

THIS NETWORK CONTAINS A TOTAL OF 58 LINKS

PEDESTRIAN DENSITY CATAGORIES

LIGHT DENSITY DENOTES A RANGE OF PEDESTRIAN VOLUME OF 0-200 PEDS./HOUR

MODERATE, 200-500 PEDS./HOUR. HEAVY, ABOVE 500 PEDS/HOUR

Figure 4. Traffic signal data.

```
INITIAL PHASE AT NODE  5 IS 1.  TIME REMAINING  0 SECONDS


NODE   5          PHASE  1          DURATION  80 SECONDS
          SIGNAL FACING LINK ( 11,  5) IS (CODE) 1
          SIGNAL FACING LINK ( 10,  5) IS (CODE) 1


INITIAL PHASE AT NODE  6 IS 1.  TIME REMAINING  0 SECONDS


NODE   6          PHASE  1          DURATION  80 SECONDS
          SIGNAL FACING LINK ( 18,  6) IS (CODE) 1
          SIGNAL FACING LINK ( 19,  6) IS (CODE) 1
```

Figure 5. Traffic demand data.

ENTRY LINK STATISTICS

LINK	FLOW RATE (VEH/HR)	PCT. TRUCKS	
(2, 10)	780	8	
(3, 11)	375	4	
(12, 11)	750	4	
(30, 29)	1110	0	
(33, 25)	1095	1	
(32, 24)	585	0	
(31, 23)	300	0	
(22, 23)	810	0	
(17, 18)	705	4	
(8, 9)	285	5	
(6, 9)	-30	0	PSEUDO-LINK

THERE ARE 10 ENTRY NODES AND 1 SOURCE/SINK NODES IN THIS NETWORK

MAXIMUM INITIALIZATION PERIOD=-240 SECONDS.

Figure 6. Link output data.

LINK	VEH-MILES	VEH DIS	MOV. TIME V-MIN	DELAY TIME V-MIN	M/T	TOTAL TIME V-MIN	T-TIME / VEH. SEC	T-TIME/ VEH-MILE SEC/MILE	D-TIME / VEH SEC	D-TIME/ VEH-MILE SEC/MILE	AVG. SPEED MPH	AVG. OCC.	STOPS /VEH	AVG SAT PCT	MSTPS /VEH	CYCL FAIL
(23, 16)	.3	19	.7	.6	.52	1.3	4.0	265.7	1.9	126.2	13.5	.3	.11	161	0.00	0
(16, 27)	2.8	25	5.4	13.6	.32	20.0	48.0	429.3	32.6	291.3	8.4	4.9	1.26	18	0.00	0
(27, 14)	.2	23	.5	.6	.41	1.1	2.9	304.9	1.7	178.6	11.8	.2	.13	134	0.00	0
(14, 18)	2.5	29	5.4	2.3	.70	7.7	15.9	187.6	4.8	56.5	19.2	1.9	.23	5	0.00	0
(18, 6)	.3	30	.6	.4	.59	1.0	2.1	228.4	.9	93.8	15.8	.2	0.00	18	0.00	0
(6, 9)	1.7	26	3.8	9.3	.29	13.1	31.4	468.9	22.3	333.7	7.7	3.3	.81	11	0.00	0
(24, 28)	4.4	36	9.1	18.2	.33	27.2	45.4	372.4	30.3	248.2	9.7	6.9	.87	14	0.00	0
(28, 19)	3.4	36	7.3	4.3	.63	11.7	19.4	206.6	7.2	76.5	17.4	2.8	.35	7	0.00	0
(19, 10)	3.1	41	6.3	13.1	.32	19.4	28.4	372.0	19.2	251.9	9.7	4.9	.62	16	0.00	0
(10, 19)	4.2	54	8.9	10.2	.47	19.1	21.2	270.9	11.3	144.2	13.3	4.8	.55	15	0.00	0
(19, 28)	4.7	50	10.7	7.0	.60	17.7	21.3	226.9	8.4	90.0	15.9	4.4	.36	11	0.00	0
(28, 24)	5.9	50	11.7	3.8	.76	15.5	18.6	158.8	4.6	38.8	22.7	3.9	.12	8	0.00	0
(11, 5)	.2	24	.5	.9	.37	1.4	3.4	370.9	2.2	233.0	9.7	.3	.13	170	0.00	0
(5, 20)	1.9	28	4.3	2.3	.65	6.6	14.1	203.6	4.9	70.7	17.7	1.6	.21	5	0.00	0
(20, 7)	.1	13	.3	.1	.79	.3	1.6	167.3	.3	34.9	21.5	.1	0.00	46	0.00	0
(7, 29)	1.1	13	2.4	.9	.72	3.4	15.5	184.3	4.3	51.1	19.5	.9	.31	2	0.00	0
(29, 15)	.1	14	.3	.2	.57	.5	2.2	235.9	.9	102.3	15.3	.1	0.00	67	0.00	0
(34, 25)	.8	16	1.9	8.4	.18	10.2	38.3	808.2	31.3	661.6	4.5	2.6	.96	12	0.00	0
(15, 34)	1.2	19	2.3	1.5	.60	3.8	12.1	199.9	4.8	79.2	18.0	1.0	.21	4	0.00	0

NETWORK STATISTICS

VEHICLE-MILES= 44.38 VEHICLE-MINUTES= 205.9 VEHICLES PROCESSED= 467 STOPS/VEHICLE= .59

MOVING/TOTAL TRIP TIME= .473 AVG. SPEED (MPH)=12.93 MEAN OCCUPANCY= 50.6 VEH. AVG DELAY/VEHICLE= 14.05 SEC

TOTAL DELAY= 108.5 MIN. DELAY/VEH-MILE= 2.44 MIN/V-MILE TRAVEL TIME/VEH-MILE= 4.64 MIN/V-MILE

Figure 7. Bus data.

LINK	NUMBER PROCESSED	MOVING TIME MIN	DELAY TIME MIN	M/T	NUMBER OF STOPS
(9, 10)	3	.0	.3	.08	2
(23, 16)	4	.2	.2	.47	1
(16, 27)	3	1.0	3.7	.22	8
(27, 14)	3	.1	.2	.24	0
(14, 18)	3	.6	1.5	.27	3
(18, 6)	3	.1	.2	.24	0
(6, 9)	3	.5	1.9	.21	4
(24, 28)	3	.5	2.5	.15	4
(28, 19)	2	.5	1.4	.27	3
(19, 10)	3	.3	1.8	.13	2

STATION	LINK	TIME CAPACITY EXCEEDED MIN	TIME EMPTY MIN	BUSES SERVICED
1	(16, 27)	0.0	2.7	4
2	(24, 28)	0.0	3.1	2
3	(16, 27)	0.0	3.0	3
4	(24, 28)	0.0	2.9	3
5	(28, 19)	0.0	3.3	2
6	(6, 9)	0.0	3.1	2

ROUTE	BUSES PROCESSED	TOTAL MOVING TIME MIN	TOTAL DELAY TIME MIN	TOTAL DWELL TIME MIN	AVG. SPEED MPH
1	2	1.3	4.6	3.6	8.3
2	2	1.1	3.6	1.1	7.6
4	3	1.8	4.6	1.6	8.1
5	3	1.5	6.9	2.9	8.9
6	0	0.0	0.0	0.0	0.0
7	1	.7	1.5	.3	11.2

SEED FOR RANDOM NUMBER GENERATOR IS 67367901

3. Traffic demand specified as flow rate (vph), percentage of trucks emitted onto the network along input (entry) links and from internal source nodes, and rate of extraction of vehicles at sink nodes;
4. Duration of simulation subintervals and specification of output options; and
5. As an option, specification of bus systems (routes, stations, mean headways, and mean dwell times) and frequency and duration of events, i.e., vehicles or conditions that block moving lanes of traffic.

MODEL OUTPUT

The user is provided considerable flexibility in specifying type and frequency of output data. All input data are printed out by the model for checking purposes. More than 50 diagnostic tests are performed, searching for inconsistencies in the input data; appropriate messages are printed identifying such errors before the program aborts. Figures 3, 4, and 5 show representative printed formats of the input data as produced by the program. Figure 6 shows cumulative statistics that are printed at the end of each simulation subinterval and more frequently if requested. The data provided for each network link and also aggregated for the entire network are as follows:

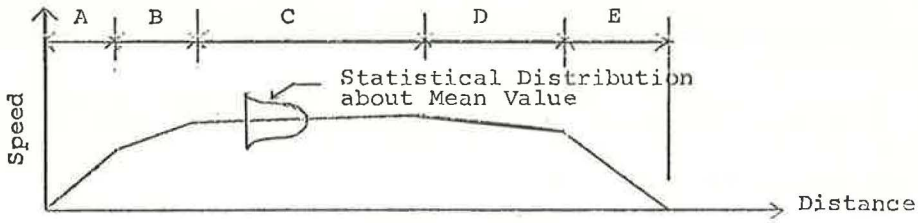
1. Link identification, by origin and destination node;
2. Estimate of total vehicle-miles of travel;
3. Count of total vehicles discharged;
4. Total vehicle moving time (at free-flow speed) in vehicle-minutes;
5. Total delay time computed as the difference between total travel time and ideal travel time based on target speed for link, in vehicle-minutes;
6. Ratio of moving time at desired speed to total travel time;
7. Total travel time in vehicle-minutes;
8. Average travel time per vehicle in seconds;
9. Average travel time per vehicle-mile in seconds per mile;
10. Average delay time per vehicle in seconds;
11. Average delay time per vehicle-mile in seconds per mile;
12. Average traffic speed in mph;
13. Average occupancy (population) in number of vehicles;
14. Percentage of vehicles stopping at least once, expressed as a decimal;
15. Average saturation percentage, expressed as the average over time of the portion of the link that is occupied by vehicles divided by its total storage capacity;
16. Total number of cycle failures, defined as the number of times queue fails to clear from the discharge end of the link during a green period; and
17. Ratio of number of vehicles stopping more than once in a link to the total number of vehicles processed.

More detailed data delineating the variation of queue length with time and signal phasing may also be (optionally) requested. If buses populate the network, the model provides detailed statistics related to bus performance, as indicated in Figure 7. Another option provides origin-destination volumes. A typical network is shown in Figure 1. An assortment of "flags" or messages focuses the user's attention on extreme conditions. For example, the message, SPILLBACK BLOCKS 2 LANES ON LINK (11, 10) AT 5 14 36, denotes a condition of excessive queuing that restricts vehicles from discharging from the afflicted link.

INTRALINK VEHICLE MOVEMENT

Because the model processes each vehicle on the network, it is possible to replicate in detail those events that occur along the streets between intersections. Hence, interaction of automobiles with buses leaving stations, impact of double-parkers, disperison of platoon, and other intralink activities may be modeled rigorously, as opposed to an idealized specification of a constant "target" speed that is unrealistic and compromises the integrity of the simulation approach.

To implement this approach, we developed a simplified car-following expression and assumed that there was an idealized form of speed profile.



The 5 phases of a typical (unimpeded) speed profile were calibrated according to the Traffic Engineering Handbook:

1. Vehicle accelerates from rest to a speed of 20 ft/sec: a value of 8 ft/sec^2 for automobiles and 3 ft/sec^2 for buses and trucks;
2. Vehicle accelerates to free-flow speed: a value of 4 ft/sec^2 for automobiles and 2 ft/sec^2 for others;
3. Unimpeded vehicle maintains stochastically assigned free-flow speed until it either responds to a lead vehicle or recognizes that it must stop;
4. Vehicle initially decelerates essentially to a nonpowered cruise of -1 ft/sec^2; and
5. Vehicle decelerates to stop at -7 ft/sec^2.

Vehicles are constrained, of course, by the trajectory of any lead vehicle. Hence, a car-following law that is superposed onto (and supersedes) the speed profile described above had to be developed. Initially, finite-difference representations of the classic differential car-following equations were considered and found wanting for this model. Instead, a model tailored to the program logic was developed, and that led to the following equation:

$$a_f = [7(s_l - s_f - V_{f1} \cdot \Delta t - L_l) + \tfrac{1}{6}(2V_l^2 - 3V_{f1}^2)]/(V_{f1} + 3) \qquad (1)$$

where

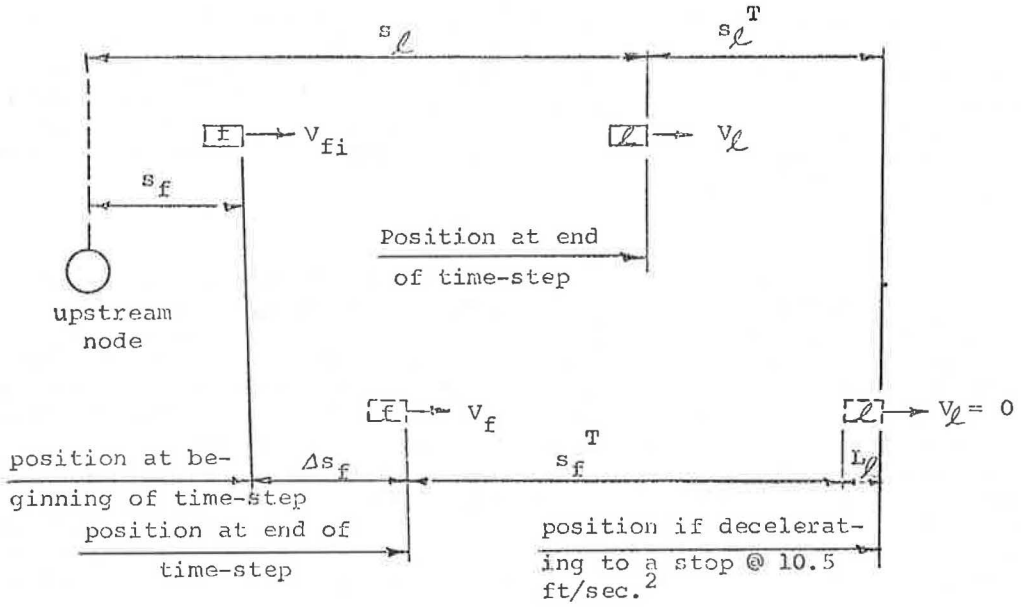
a_f = acceleration of the following vehicle now being processed;
Δt = simulation time step = 1 sec;
V_{f1} = speed of following vehicle at beginning of this time step;
V_l = speed of lead vehicle at end of this time step;
L_l = effective length of lead vehicle;
s_l = distance of lead vehicle from upstream node at end of time step; and
s_f = distance of following vehicle from upstream node at beginning of time step.

This expression is derived under the assumption that a "normal" deceleration is 7 ft/sec^2; the application of this value and assignment of $\Delta t = 1$ account for the apparent inconsistency in units. No attempt was made to validate this expression rigorously; detailed testing uncovered no instabilities in vehicle trajectories and indicated realistic vehicle headways.

In the following sketch,

s_l^I = distance traveled by the lead vehicle during deceleration at 10.5 ft/sec^2 (50 percent above "normal") to a stopped position; and
s_f^I = distance traveled by following vehicle during deceleration at 7 ft/sec^2 to a stopped position.

The model seeks the value of acceleration (or deceleration) that the following vehicle must now (at the beginning of the time step) apply to prevent a collision with the lead vehicle, if the latter should begin to decelerate at 10.5 ft/sec^2 to a stop. The following vehicle's deceleration, if this event should occur, would be normal, i.e., 7 ft/sec^2. The rationale is that a driver anticipates a sudden deceleration (50 percent above normal) and maintains his speed and spacing appropriately to prevent a collision. This



scenario may be written mathematically (in the limiting case of near collision) as

$$s_f^I = s_\ell + s_\ell^I - s_f - \Delta s_f - L_\ell \qquad (2)$$

Writing

$s_\ell^I = V_\ell^2/2d_\ell = V_\ell^2/21,$

$s_f^I = V_f^2/14,$

$V_f = V_{f1} + \Delta V,$

$V_f^2 \rightarrow V_{f1}^2 + 2V_{f1}\,\Delta V,$

$\Delta s_f \rightarrow \frac{1}{2}(V_{f1} + V_f)\,\Delta t,$

$V = a_f \cdot \Delta t,$

$\Delta t = 1,$

and substituting appropriately into Eq. 2 yield

$$(V_{f1}^2 + 2V_{f1}a_f)/14 = s_\ell + (V_\ell^2/21) - s_f - [V_{f1} + (a_f/2)] - L_\ell$$

Solving for a_f yields the expression shown above (½ rounded to 3).

QUEUE DISCHARGE

Each vehicle at the head of a queue prior to the onset of green is stochastically assigned a start-up delay that must be exhausted after the green phase is activated and before it is discharged. Each following vehicle in the queue is stochastically assigned a discharge headway that is related to the specified mean headway value, the statistical distribution about that mean, and its original position in the queue. When a vehicle is discharged, the remaining members of the queue move up in response to a "green wave" propagating upstream at a speed of 1 vehicle (≈20 ft) per sec. Hence, the ninth vehicle in a queue when the green phase is activated remains motionless for 8 sec after the first vehicle discharges.

Left-turn movements represent a critical component of design. For unprotected phases, a left-turn vehicle first in queue may "jump the gun" before oncoming traffic (determined probabalistically), seek an acceptable gap in the oncoming traffic (also determined probabalistically), or negotiate the turn during the following amber signal. A multitude of tests are made for each vehicle ready to discharge, depending on its intended maneuver, type, status of traffic on receiving link, its lane assignment in that link, and many other factors.

BUS TRAFFIC

It is well recognized that, although urban bus vehicles constitute a relatively small percentage of total traffic, their impact on general operations is profound. The need for buses to maneuver at stations, the blockage effect of buses dwelling at stations, their size, and their sluggish operating characteristics, all contribute to degrading overall traffic performance. Considerable effort was expended to ensure that bus traffic was explicitly and realistically treated by the model. Vehicles, identified as buses in the sense of vehicle length and acceleration characteristics, traverse prescribed paths (routes) through the network servicing those stations assigned to that route. The probability of stopping and the duration of dwell are assigned to each vehicle stochastically. Impedance with other traffic, queuing that prevents buses from accessing their stations, possibility of station storage being saturated, and many similar factors are rigorously modeled. Separate sets of statistics are accumulated and displayed for bus traffic.

RESPONSIVE SIGNAL CONTROL

The model was designed for the express purpose of evaluating responsive signal control policies. To this end, the ability of the model to identify the status of each vehicle—lane position, location, speed, type, and intended maneuver—provides it with the means of replicating almost any control policy, either local (intersection-specific) or global (system-wide). Such policies may be introduced into the model, through the medium of additional FORTRAN code, with no interface problems whatever. This is achieved by providing as many as 9 "windows" in the code that permit the addition of subroutines to reflect these policies. Each subroutine, representing a particular policy, may refer to one or more specified intersections of the network. In addition, a computer-monitored control system, responding to signals accessed by detectors as determined by the surveillance simulation (another optional component of the model), can be rigorously simulated by UTCS-1 to predict its overall performance prior to implementation.

REFERENCE

1. Bruggeman, J., Lieberman, E., and Worrall, R., Network Flow Simulation for Urban Traffic Control System. KLD Assoc., Inc., Tech. Rept. FH-11-7462-2, 1971.

GAP-ACCEPTANCE CHARACTERISTICS IN FREEWAY TRAFFIC FLOW

Juergen Pahl, University of California, Los Angeles

Aerial data were used to determine average spatial and time sizes of the accepted gaps and the lag gaps in gap-acceptance maneuvers for exiting vehicles close to their intended off-ramp and for through vehicles as a function of distance from the off-ramp. Furthermore, an attempt was made to quantify the accident risk that was assumed by each lane changer and the accident risk that was imposed by each lane changer on the lag car of the accepted gap. Results were obtained for an 8-, a 6-, and a 4-lane freeway site and for various flow levels. In general, the average spatial and time gap and lag gap sizes of gaps accepted by exiting vehicles decrease as the exiting vehicles approach the off-ramp. The corresponding average values for through vehicles show less, if any, dependency on distance from the off-ramp. Average gap and lag gap sizes of gaps accepted by exiting and through vehicles decrease with increasing flow levels. Within the assumptions made, it was found that a larger accident risk is accepted by exiting vehicles than by through vehicles in their lane changes, whereas the accident risk imposed on the lag car of the accepted gap is smaller for lane changes by exiting vehicle than by through vehicle.

•THE AVAILABILITY of quantitative data for gap-acceptance characteristics is very important for the understanding of quantitative features of freeway traffic flow. Some rather general results on gap-acceptance characteristics are available (1, 2, 3), but more detailed results are urgently needed. Recently, a methodology was developed (4) that enabled the acquisition of data suitable for this purpose. This methodology involves the application of aerial photographic techniques to obtaining discrete trajectories of vehicles traversing a given freeway site and the subsequent analysis of these trajectories to discern quantitative features of the freeway traffic flow.

In this paper, gap-acceptance characteristics are investigated for exiting vehicles close to their intended off-ramps and for through vehicles. The investigated parameters include spatial and time sizes of the accepted gaps and of the lag gaps in each gap-acceptance maneuver. For each gap-acceptance maneuver, a further attempt was made to quantify the accident risk that was assumed by the lane changer as well as the accident risk that was imposed by the lane changer on the follower of the accepted gap.

DATA

The 3 freeway sites selected each contained an off-ramp and an upstream section of the freeway. The sites were situated in the Los Angeles area and consisted of an 8-lane freeway (Ventura Freeway, westbound, at White Oak Avenue off-ramp), a 6-lane freeway (Santa Ana Freeway, southbound, at Washington Boulevard off-ramp), and a 4-lane freeway (Newport Freeway, northbound, at Katella Avenue off-ramp).

The data were collected with a 70-mm Maurer camera operated from a helicopter that hovered over the studied freeway site. The camera took pictures at time intervals of 1 sec, and the length of the photographed freeway section was of the order of 1 mile. The pictures were scanned in successive order so as to obtain digitized positions of vehicles. A specialized software system deduced individual trajectories from the digitized vehicle coordinates. The resulting trajectories yield the space-time history of each vehicle within the studied freeway site. They include longitudinal and lateral position and speed for each car in time intervals of 1 sec.

The acquired data were subjected to a statistical analysis to isolate those time periods in which the average flow level remained constant. The length of the obtained time periods ranged from 1.5 to 15 min. The data were aggregated into groups where the average flow levels of the selected time periods differ only within a small range in order to improve the statistical significance of the derived analysis results. Table 1 gives the resulting data groups that were used for this study (more details are given in another report, 4).

SPATIAL AND TIME CHARACTERISTICS OF GAP-ACCEPTANCE MANEUVERS

In this study of gap-acceptance characteristics as a function of distance from the intended off-ramp of exiting vehicles, each study site was divided into zones of 600-ft lengths starting at the nose of the off-ramp and progressing upstream. In each zone all lane changes of exiting and through vehicles were determined, and for each lane change the leader and the follower of the gap that was accepted by the lane changer were defined. An accepted gap was always assigned to the same zone in which the lane change occurred, even though the leader or follower of the gap or both may have been positioned outside this zone at the moment of the lane change. Figure 1 shows a sketch of a gap-acceptance maneuver and some of the parameters of interest in such a maneuver.

For each zone, the spatial sizes of all accepted gaps, $s_{g,i}$ (front bumper to front bumper), were determined at the moment of the lane change. For through vehicles and exiting vehicles separately, the gap sizes were classified into 50-ft intervals, and the sample probability density

$$n_k(d)/[N(d) \times 50 \text{ ft}] \tag{1}$$

was computed for each zone, where $n_k(d)$ denotes the number of gaps in the kth gap-size interval in the zone at distance d from the off-ramp, and N(d) denotes the total number of lane changes of through and exiting vehicles respectively in the same zone. For each zone, the estimators of the average gap size

$$\bar{s}_g = (1/N) \sum_{i=1}^{N} s_{g,i} \tag{2}$$

and its variance

$$\sigma_g^2 = 1/\ N(N-1) \sum_{i=1}^{N} (s_{g,i} - \bar{s}_g)^2 \tag{3}$$

were also computed.

Analogous procedures were applied to all lag gap sizes, $s_{f,i}$.

In addition to this study of spatial gap and lag gap sizes, an analogous study for the associated time gap and lag gap sizes was also performed. For this purpose, the time size of each accepted gap was computed as

$$h_g = s_g/v_f \tag{4}$$

and the time size of the lag gap was computed as

$$h_f = (s_a/v_f) - (s_c/v_c) \tag{5}$$

The meaning of the symbols is shown in Figure 1. For through vehicles and exiting vehicles separately, these time gaps and lag gaps were classified into 0.5-sec intervals, and the sample probability density was computed as

Table 1. Average flow levels of study sites.

Number of Freeway Lanes	Flow	Avg Number of Vehicles/Hour		Time Period (sec)
		Through	Exiting	
8	Low	4,450	437	1,040
	Medium	5,520	595	1,120
	High	7,320	771	2,525
6	Medium	4,440	484	400
	High	5,680	560	180
4	Medium	2,520	427	1,405

Figure 1. Gap-acceptance maneuver.

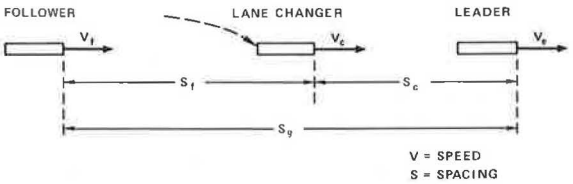


Figure 2. Spatial size of gaps accepted by exiting vehicles, 8-lane site, high flow.

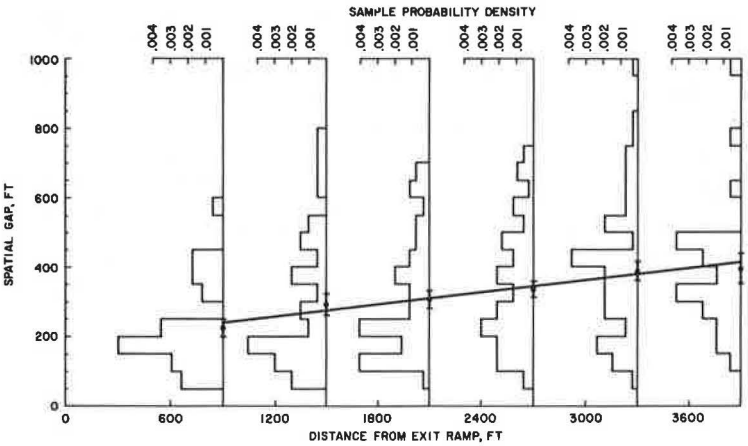
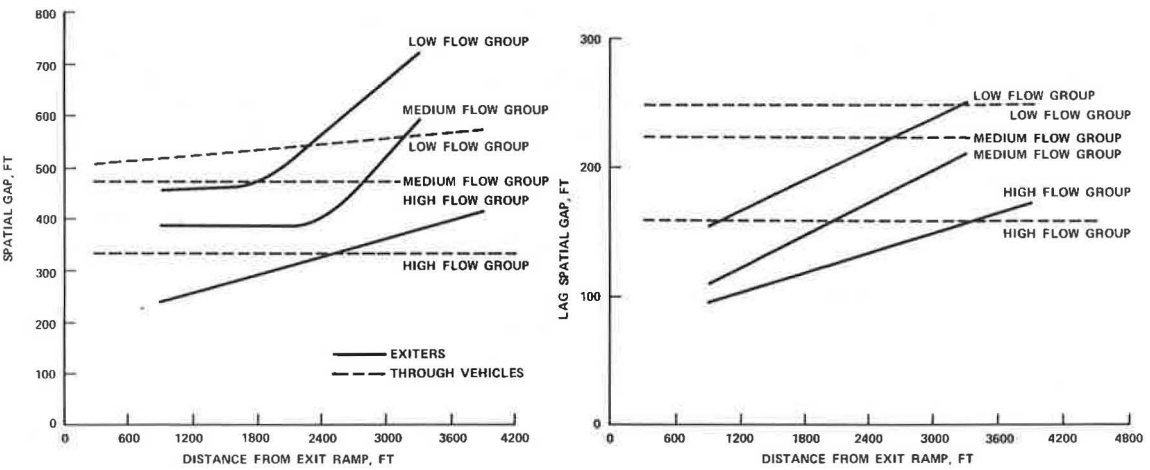


Figure 3. Estimated average spatial gaps accepted, 8-lane site.



$$n_k'(d)/[N(d) \times 0.5 \text{ sec}] \quad (6)$$

for each zone, where $n_k'(d)$ denotes the number of gaps in the kth time gap size interval for through and exiting vehicles respectively in the zone at distance d from the nose of the off-ramp. Average time gap sizes and their variances were estimated by equations similar to Eqs. 2 and 3. Analogous results were computed for the time sizes of the lag gaps.

A detailed graphical presentation of the obtained results is given in another report (4, Appendix D). Figure 2 shows an example that is typical of the results obtained for the spatial sizes of gaps accepted by exiting vehicles in the high-flow group on the 8-lane freeway. So that the obtained results could be easily compared, smooth curves were fitted through the average values obtained in each data group, and these curves are shown in Figures 3 through 8.

ACCIDENT RISK

An attempt was also made to quantify the accident risk involved in the gap-acceptance maneuvers. For this purpose, a computation was made of how long the follower of a gap could delay his braking onset after the leader of the gap had started to brake, such that the follower just barely avoided a collision (this is a modification of the St. John and Kobett time sum concept, 5). In the following these results will be called the "permissible braking delay." For example, the distance between the front bumper of the lane changer and the back bumper of the leader (Fig. 1) is s_c - 17 ft, based on the generally used value of 17 ft for the average car length. It is assumed that the leader starts to brake at time t = 0 with a constant deceleration $\ddot{x}$. It is further assumed that the lane changer will decelerate with the same deceleration $\ddot{x}$ when his brakes are activated. In order to just barely avoid a collision with the leader, the lane changer has then the permissible braking delay

$$T_c = [(s_c - 17 \text{ ft})/v_c] - (1/2v_c\ddot{x})(v_l^2 - v_c^2) \quad (7)$$

The permissible braking delay for the follower (Fig. 1) with respect to the lane changer is

$$T_f = [(s_f - 17 \text{ ft})/v_f] - (1/2v_f\ddot{x})(v_c^2 - v_f^2) \quad (8)$$

For this study, the deceleration value, $\ddot{x} = -10$ ft/sec^2, was chosen for the computations; this value corresponds to an intermediate risk level (5).

Permissible braking delays ranged typically from values as low as -3 sec to values of more than 10 sec. An earlier report (4) showed the sample probability densities of the permissible braking delays for the various data groups as a function of distance from the off-ramp. Figure 9 shows a typical example that was obtained for exiting-vehicle lane changes in the high-flow group on the 8-lane freeway site.

Negative values of permissible braking delays indicate very high risk levels because a collision is unavoidable, under the above assumptions, if the leader or the lane changer were suddenly to decelerate 10 ft/sec^2. However, because typical reaction times are of the order of 1 sec, even all permissible braking delays, $T_c \lesssim 1$ sec and $T_f \lesssim 1$ sec, indicate high risk levels.

With this assumption, the percentage of cars with high accident risks in gap-acceptance maneuvers were obtained for each data group. The results are given in Table 2. The dependency of these percentages on distance from the off-ramp could not be studied because the available data samples were not large enough for this purpose. Because the available sample sizes varied considerably among the various data groups, the accuracies of the percentages given in Table 2 are quite different from one another. A rough indication of the statistical significance of each percentage given is provided by the total number of lane changes from which each percentage was computed.

SUMMARY AND CONCLUSIONS

Figures 3 through 8 show that the average spatial and time gap and lag gap sizes of gaps accepted by exiting and through vehicles decrease with increasing flow levels.

Figure 4. Estimated average time gaps accepted, 8-lane site.

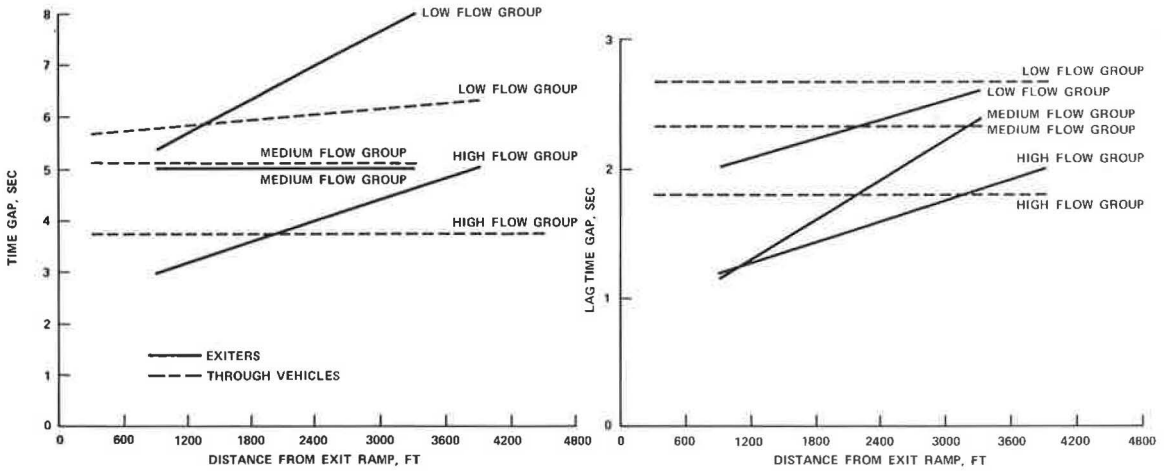


Figure 5. Estimated average spatial gaps accepted by through vehicles, 6-lane site.

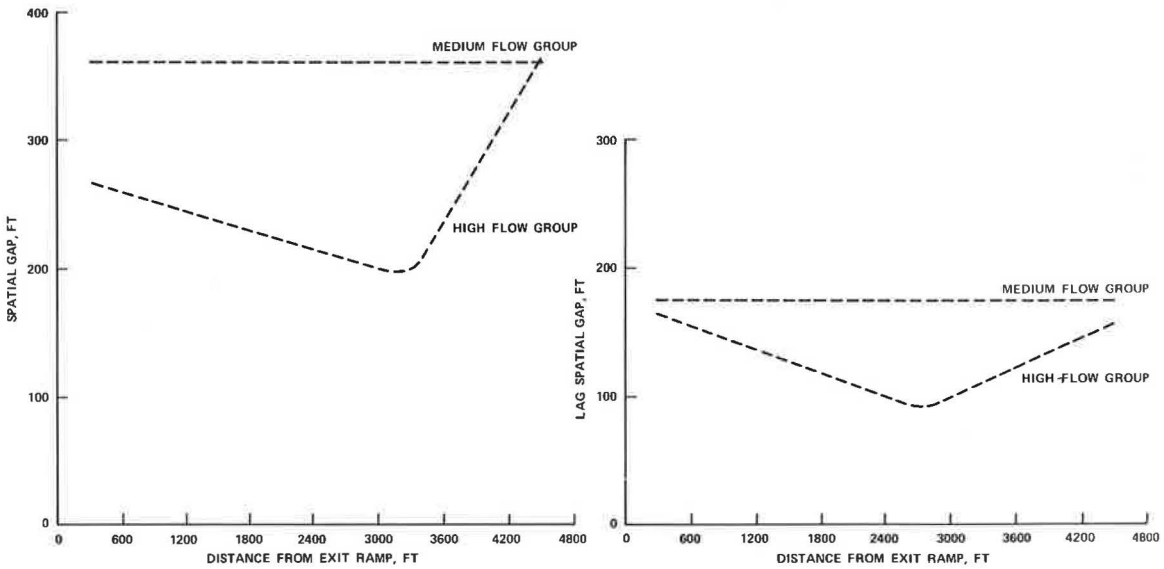


Figure 6. Estimated average time gaps accepted by through vehicles, 6-lane site.

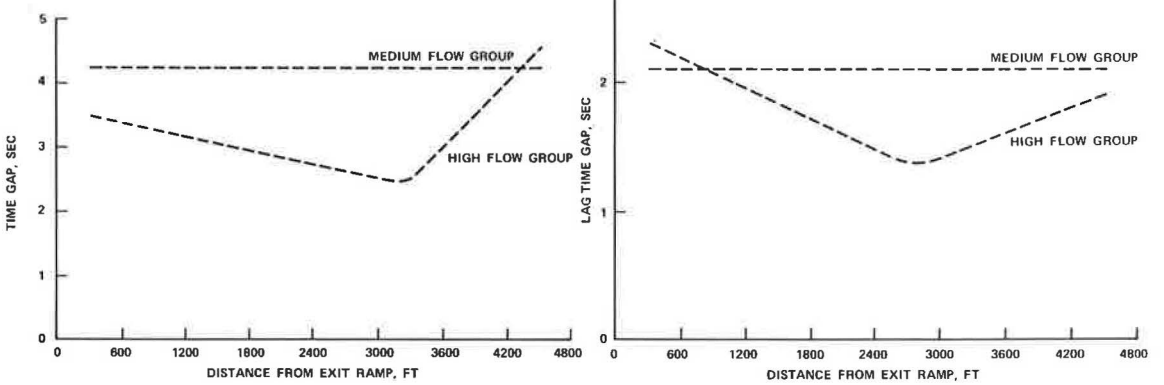


Figure 7. Estimate average spatial gaps accepted, 4-lane site, medium flow.

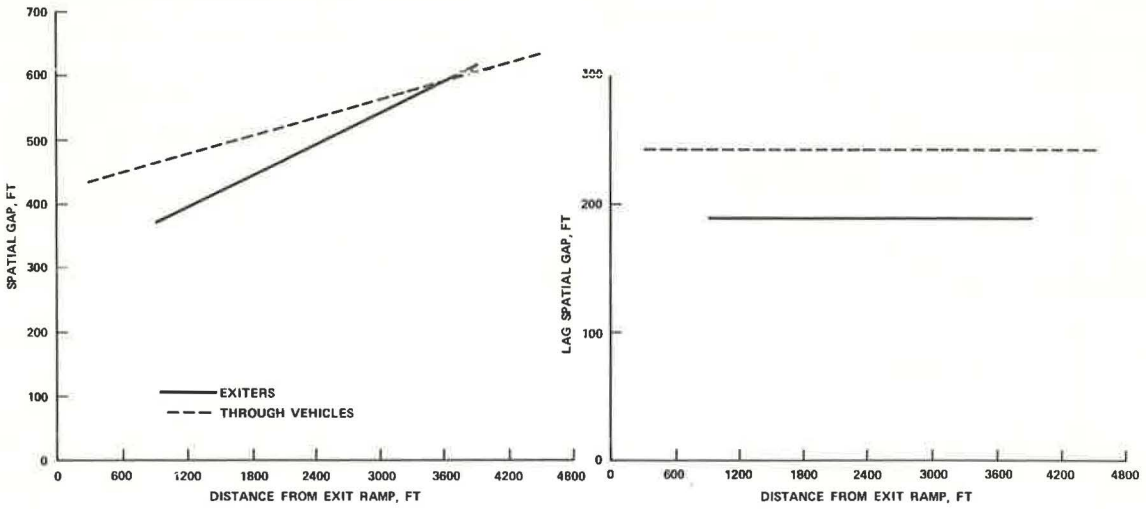


Figure 8. Estimated average time gaps accepted, 4-lane site, medium flow.

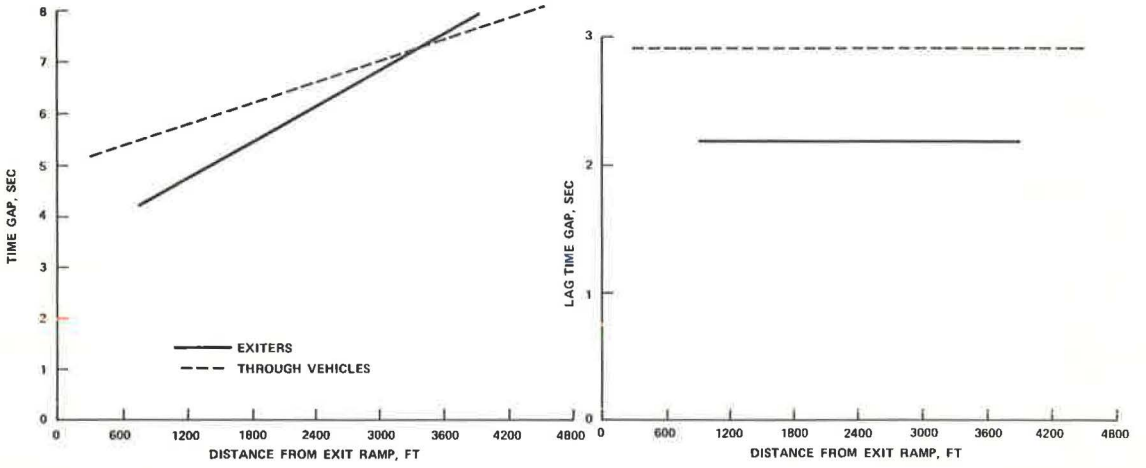


Figure 9. Permissible braking delay of exiting vehicles when accepting gap, 8-lane site, high flow.

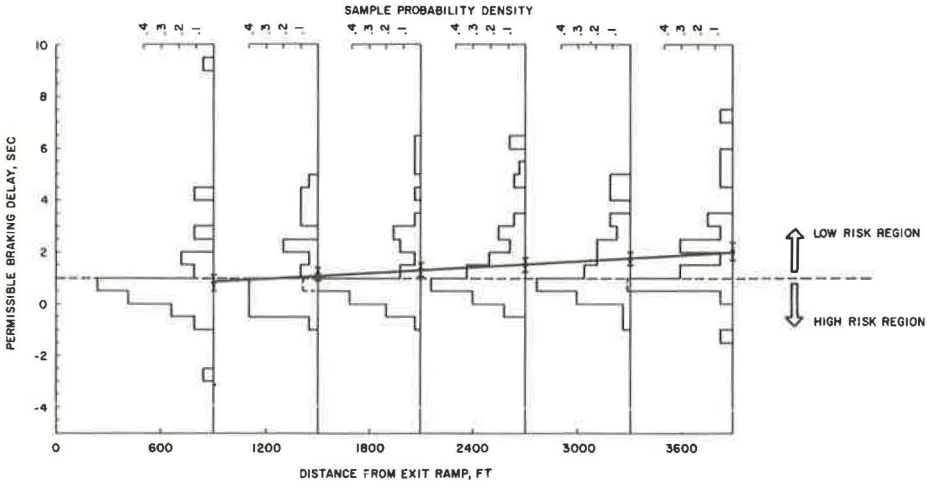


Table 2. Permissible braking delays less than 1 sec for lane-changing vehicles and for lag cars.

Number of Lanes and Flow	Vehicles Changing Lanes				Vehicles Accepting Gaps of Lag Cars			
	Exiting		Through		Exiting		Through	
	Percent	Number	Percent	Number	Percent	Number	Percent	Number
8, low	34	87	22	379	9	87	18	379
8, medium	32	75	22	500	21	75	23	500
8, high	55	271	29	1,399	19	271	34	1,399
6, medium	31	26	16	93	12	26	34	93
6, high	86	7	16	50	29	7	46	50
4, medium	30	50	14	246	14	50	22	246

Figures 3 and 4 show that, for the 8-lane freeway, the spatial and the time average sizes of the gaps accepted by exiting vehicles decrease with decreasing distance from the off-ramp. The only possible exception to this general trend may occur for the time sizes of gaps accepted by exiting vehicles in the medium-flow group; but here, the statistical accuracy of the data points is especially poor (4). Figures 3 and 4 also show that the corresponding spatial and time lags in the gap-acceptance maneuvers of exiting vehicles exhibit also a substantial decrease with decreasing distance from the off-ramp.

For the 4-lane freeway, Figures 7 and 8 show a decrease in spatial and time average size of gaps accepted by exiting vehicles for decreasing distances from the off-ramp. However, in this case, the spatial and time lag in gap-acceptance maneuvers of exiting vehicles does not change with distance from the off-ramp. Unfortunately, because of a lack of a sufficiently large data sample, no graphs are available for gap-acceptance maneuvers of exiting vehicles on the 6-lane freeway site.

The corresponding average values for through vehicles show less, if any, dependency on distance from the off-ramp. Figures 7 and 8 show that through vehicles accept, on the average, shorter gaps with decreasing distance from the off-ramp for the medium-flow group of the 4-lane site. Figures 3 and 4 show a similar, although weak, trend in the low-flow group of the 8-lane site. No such trend is apparent in the medium- and high-flow groups of both the 8-lane and the 6-lane sites. The spatial and time lags in gap-acceptance maneuvers of through vehicles do not depend, on the average, significantly on distance from the off-ramp for any of the data groups, with the exception of the high-flow group of the 6-lane freeway. Figures 5 and 6 also show that, for the high-flow group of the 6-lane freeway, the average values for through vehicles exhibit a special and unique dependency on distance from the off-ramp.

Data given in Table 2 show that, as expected, a substantially larger accident risk is accepted by exiting vehicles than by through vehicles in their lane changes. Furthermore, the accident risk of exiting vehicles in their lane changes tends to increase with increasing flow levels. Somewhat unexpected, however, is the finding that the accident risk imposed on the lag car of the accepted gap is substantially smaller for exiting-vehicle lane changes than for through-vehicle lane changes.

REFERENCES

1. Worrall, R. D., and Bullen, A. G. R. An Empirical Analysis of Lane Changing on Multilane Highways. Highway Research Record 303, 1970, pp. 30-43.
2. Drew, D. R. Gap Acceptance Characteristics for Ramp-Freeway Surveillance and Control. Highway Research Record 157, 1967, pp. 108-143.
3. Wynn, F. H. Weaving Practices on One-Way Highways. Bureau of Highway Traffic, Yale Univ., May 1946, pp. 9-57.
4. Pahl, J., Charles, S. E., and Knobel, H. C. Exit Ramp Effects on Freeway System Operation and Control. Univ. of California, Los Angeles, Eng. Rept. 71-49, 1971.
5. St. John, A. D., and Kobett, D. R. Traffic Simulation for the Design of Uniform Service Roads in Mountainous Terrain. Midwest Research Institute, Final Rept., Vols. 1 and 2, 1970.

COMPARATIVE STUDY OF TRAFFIC CONTROL CONCEPTS AND ALGORITHMS

P. K. Munjal and Y. S. Hsu, System Development Corporation,
Santa Monica, California

Traffic network control has been under extensive research in recent years. Efforts and results are diversified in many respects, mainly because of the various mathematical models selected to describe the network, the different prediction techniques, if any, and the stated objectives. In this paper we discuss the merits and shortcomings of traffic control concepts and algorithms that have appeared in major publications. Starting with an isolated intersection and continuing to arterials and networks, all discussions are in terms of the mathematical model, the prediction, and the optimization of objectives. Some current network control programs in progress are also addressed.

•THE CONCEPT of controlling traffic operations in a network is currently undergoing extensive research. For purposes of discussion, the majority of notable results appearing in the literature can be grouped into the following important components: the mathematical model, the prediction, and the objective function.

A mathematical model is a set of mathematical expressions or equations that represent traffic dynamics and control. Factors influencing most mathematical models in this field include

1. The configuration of the network, i.e., an isolated intersection, an arterial, a tree type of network, or a general network including closed loops; and
2. The behavior of traffic, i.e., uniform, Poisson, binomial, or any other well-defined probability distribution of vehicle arrivals and constant or random platoon density.

Although a more sophisticated model is likely to produce more accurate results, the computational load is also likely to increase.

Prediction is the forecast of traffic flow patterns for the near future, based on historical data and current observations.

Objective function refers to the specific objective of traffic control. In the signal setting of an isolated intersection, the objective is usually to minimize the average delay per car. In that for an arterial, the maximization of bandwidth is usually cited as the objective. In control of a network, a more general objective function—minimizing a weighted sum of delay and stops—is gradually becoming a universal criterion.

This paper presents a broad investigation of important traffic control concepts and algorithms from selected, recently published articles. Far from being all-inclusive, our purpose is to illustrate the important features of the subject and to cover as wide a spectrum as possible within the limited scope of this paper. Although our focus is primarily on sophistication and applicability, we shall discuss the concepts in terms of the models used; the prediction techniques employed, if any; and the defined objective functions.

MATHEMATICAL MODEL

The simplest model is one that describes an isolated intersection. An intersection is isolated if arrivals in each approach can be considered as a population drawn independently from some well-defined probability distributions. A uniform model, first analyzed by Clayton (8), is the simplest car-arrival model of an isolated intersection.

A model is uniform if the headway has a constant mean and zero variance. A more complicated car-arrival model is the binomial arrival model where cars arrive independently from a binomial distribution. This model was first analyzed by Winsten (6). The well-known work by Webster (47) makes the assumption that arrivals form a Poisson process. Newell (30) used a model in which the arrival headway has a shifted exponential density

$$f(t) = \lambda e^{-\lambda(t-t_0)} \qquad t \geq t_0$$

$$f(t) = 0 \qquad t < t_0 \tag{1}$$

where $t_0 = 1/s$, s is the saturation flow rate, and λ is the flow rate. A truncated exponential function

$$f(t) = \lambda e^{-\lambda t}/(1 - e^{-4}) \qquad 0 \leq t \leq 4/\lambda \tag{2}$$

was proposed by Gafarian et al. (9). A combination of the 2 models given above used by Munjal and Fitzgerald (27) has the following form:

$$f(t) = [\lambda/(1 - e^{-4})]e^{-\lambda(t-t_0)} \qquad t_0 \leq t \leq 4/\lambda + t_0$$

$$f(t) = 0 \qquad \text{otherwise} \tag{3}$$

This is called a truncated-shifted exponential density function.

The Poisson and exponential type of distributions are among the most widely used. The nonshifted exponential distribution is inferior to other distributions in single-lane traffic because it has the maximum probability density at zero headway, which is contrary to our experience that the probability density is zero at zero headway, gradually increases to a maximum at some point depending on traffic demand, and then decreases as headway increases. In this respect, it seems a Poisson distribution is more realistic in terms of describing traffic. Similarly, we could also devise a shifted Poisson distribution function like the shifted exponential. It is also noted that when traffic is very light vehicles can overtake freely, and car arrivals are independent. When traffic becomes heavy or the overtaking is limited by the environment, then some drivers, on catching up to slower vehicles, are forced to follow behind. In this circumstance, a simple independence of individual vehicle can be substituted by a more sophisticated independence of queues. Miller (24) and Tanner (39) have analyzed this problem, and their results are in a more complicated mathematical form than those described above.

Many other arrival models have also been presented in the literature. Each model results in a different expression of delay—and, therefore, in a different optimal cycle time—for the objective of minimizing delay.

Seldom are intersections in urban areas isolated from one another. In other words, vehicles do not arrive independently but rather in platoons. The simplest model to describe the platoon is the uniform model, which assumes that the platoon has a constant density and travels at a constant speed. The work of Morgan and Little (26), Little (20), and Yardeni (50) on signal settings of an arterial is based on the uniformity of platoons. For networks, SIGOP (40), the combination method by Hillier (16), and the work of Inose et al. (19) and Chang (7) all use the assumption of uniform platoons. Because of its simplicity, the uniform model is likely to be inaccurate and unrealistic.

An improvement of the uniform model is to consider the deceleration and compression of the platoon when it stops for a red signal and the acceleration and dispersion of the platoon when it starts moving at the beginning of a green signal. In the Sperry Rand study (38), the calculations for queue, delay, and travel time are based on the uniform platoon assumption and a set of experimentally obtained acceleration and deceleration characteristics.

The TRANSYT model by Robertson (36) is a more sophisticated model that uses experimentally derived 50-part histograms to simulate downstream platoon shapes. The model is claimed to be significantly more accurate than uniform models, for it counts the random fluctuations of the platoons. The work by Allsop (2) illustrates the relation between platoon shape and traveling distance from a signal light. Other models that consider a single intersection linked to nearby intersections include Miller (25) and Weinberg et al. (48). They essentially use detector-measured data and prediction techniques to relate the delay, traffic movements, and signalization.

For control purposes, we like to have a sophisticated model that provides accurate information for prediction and optimization. However, this is likely to increase the model's complexity, which makes optimization a much more difficult task. The need, at present, is to have for a general network a modular mathematical model that, when transformed into a digital computer simulation model, will have high real-time/simulated real-time ratio and can provide a test bed for any control algorithms before they are applied to a real traffic network.

PREDICTION

Most conventional control techniques do not require prediction. Signal settings are either completely determined by detector measurements—as in the case of fully or semiactuated controllers—or predetermined, based on historical and time-of-day traffic data—as in the case of multidial systems or of a malfunctioning detector. The more advanced traffic control techniques or traffic-responsive controls do require that certain traffic data be predicted for later use in the optimization stage. The amount of data to be predicted and the method used for prediction are certainly questions interwoven with the optimization problem and the mathematical model.

If an intersection is isolated and arrival is uniform, it is clear that no prediction is needed. The best signal setting is one that allows just sufficient effective green time to each approach, provided the traffic is unsaturated. Unfortunately, this is not the case in real traffic. In the work by Miller (24) and Weinberg et al. (48), exponentially weighted, moving-average predictors have been employed for estimating car arrivals during the next few seconds. This type of predictor has the general form

$$\hat{X}_{n+1} = X_n + \lambda(\hat{X}_n - X_n) \tag{4}$$

where $\hat{X}_{n+1}$ is the predicted value for time $n + 1$, X_n is the measured value at time n, $\hat{X}_n$ is the past prediction for time n, and λ is a positive constant less than 1. This method is fairly simple, but its statistical properties and how to make the best choice of λ are still unknown. In TRANSYT (36), an expression given for λ is claimed to best fit the data.

For fast-changing traffic, the moving-average type of predictors will produce considerable lag in the predictor. An improvement could be made by a consideration of the increasing rate or decreasing rate. That is, instead of Eq. 4, we have

$$\hat{X}_{n+1} = X_n + \lambda(\hat{X}_n - X_n) + \Delta_n$$

$$\Delta_n = \gamma(X_n - X_{n-1}) + (1 - \gamma)\Delta_{n-1} \tag{5}$$

for $0 \leq \gamma \leq 1$ and some initial value of Δ_1 depending on the traffic pattern. The recursive predictor Eq. 5 is believed capable of catching the trend of traffic more quickly.

TRW is currently conducting a network control software research program that requires an accurate prediction of traffic patterns 15 min into the future. It is believed that any dynamic type of control (real-time or variable cycle timing) requires fast and accurate prediction of traffic flow patterns. Here we have a dilemma: If we predict traffic patterns farther into the future, the optimization process is able to employ more future information and is likely to produce better signal settings; on the other hand,

the predicted value is likely to reflect greater error. A good procedure seems to be the following:

1. Use all currently available information to form predictions;
2. Input the predicted traffic load to the mathematical model and optimize its output (minimizing delays, stops, and the like) for some fairly long time period in the future, say, T min;
3. Follow this optimal strategy for a short time period, say, ΔT min, $\Delta T < T$;
4. Update the predictions based on the new information; and
5. Recompute the optimal strategy and repeat the process.

This procedure, in fact, illustrates the interrelation of the mathematical model, the prediction, and the optimization.

OPTIMIZATION OF OBJECTIVE FUNCTIONS

In conventional traffic control methods, e.g., traffic-actuated or 3-dial systems, signal settings are determined either by trial and error or by engineering judgment. The performance is difficult to evaluate by any criterion except "it is known to produce satisfactory results under certain conditions."

In this section we shall discuss systematic methods of signal settings. Results found in current literature fall into 3 categories of control: a single intersection, an arterial, and a network. We shall examine each category.

Signal Settings of a Single Intersection

As previously stated, the simplest model of a single intersection assumes that arrivals form an independent and identically distributed random sequence for each approach. In this case, the departure rate is either zero when the signal is red (or, rather, effective red, when we count the lost time), or the saturation flow rate when queue has not dissipated completely, or merely the arrival rate when there are no queues in that approach when the signal is in its effective green duration. Different situations arise when the single intersection model is treated deterministically or stochastically. We shall examine these 2 cases separately.

There are, in general, 2 approaches for controlling a 2-way intersection: the vehicle-actuated control and the fixed-cycle control. Conventional vehicle-actuated controllers, which include semiactuated and fully actuated controllers, yield good results when traffic is light but fail to operate properly when traffic is heavy. They are also observed to yield poor results for linked intersections. More advanced vehicle-actuated controllers use computers in which a set of algorithms is employed to obtain signal settings based on currently measured and predicted traffic data. The works of Miller (25), Weinberg et al. (48), Grafton and Newell (14), and Martin-Löf (22) belong to this category. In the last two, a dynamic programming approach is used to obtain optimal signal settings. As noted by Newell (33) and Newell and Osuna (34), a general rule that usually yields minimum delay is to switch the signal as soon as the queue is dissipated. For this type of controller, there is no cycle time; the minimum green, the maximum green, and the switch-command signal determine the signal settings.

For fixed-cycle signals, Clayton (8) was probably the first to study the behavior of an isolated intersection analytically. His work was followed by that of Wardrop (44). Both assumed stationary, deterministic arrival and departure in unsaturated traffic. Total delay is minimized by the optimal cycle time being just long enough in each approach to discharge all vehicles that arrive in that approach during the cycle.

In general, it can be shown that the average delay time, D_i, per car for approach i for any intersection (not necessarily an intersection with 2 one-way streets) is

$$D_i = r_i^2/\{2C[1 - (d_i/s_i)]\} \quad (6)$$

where d_i, s_i, and r_i = split, arrival rate, and effective red time for approach i; and C = cycle time.

As we shall see later, the expression for D_i in Eq. 6 often appears as the first term in the expression for expected delay per car in a stochastic model.

Although it is a simple matter to obtain optimal cycle time for regular arrivals, the result may not have wide application because of the random behavior of traffic. Another widely studied model is the Poisson arrival model. The first and most extensively used expression for delay of a Poisson model for any intersection is due to Webster (47). This is,

$$D_i = \{[C(1 - \lambda)^2]/[2(1 - \lambda X)]\} + \{X^2/[2d_i(1 - X)]\} - 0.65(C/d_i^2)^{1/3}X^{2+5\lambda} \quad (7)$$

where

$\lambda = 1 - (r_i/C)$, ratio of effective green to cycle time for approach i;
d_i = average arrival rate for approach i; and
$\lambda X = d_i/s_i$, and X is degree of saturation.

The first term in Eq. 7 is the same as the term in Eq. 6. The additional delay caused by Poisson arrival is due to the fact that a queue is not necessarily cleared during each cycle. To minimize Eq. 7 Webster found the best cycle to be approximately

$$C = (1.5L + 5)/[1 - \Sigma(d_i/s_i)]\text{sec} \quad (8)$$

where $L = \Sigma L_i$, and L_i is the lost time for approach i. Therefore, the cycle times are higher for random arrivals than for regular arrivals. When the optimal cycle is determined, the choice of split is a simple matter. This is given by

$$g_i = (d_i/s_i)\,C + 1.5L_i + (5/n) \quad (9)$$

for i = 1, 2, ..., n, and for n-phase signal. It is easy to verify that

$$\sum_{i=1}^{n} g_i = C \quad (10)$$

The discussions given above for the determination of optimal signal settings are only for unsaturated intersections, that is, when the denominator in Eq. 8 is greater than zero. When an intersection frequently saturates or oversaturates, i.e., when the denominator is equal to or less than zero, it is called a critical intersection. Gazis and Potts were probably the first to consider the oversaturated intersection, specifically an isolated intersection with 1-way uniform traffic flows in which the demand is a linear and deterministic function with respect to time. They observed that the cycle time and split to minimize the total delay are those that dissolve both queues simultaneously at the end of oversaturation. In a later paper, Gazis (12) considered 2 interconnected intersections containing 3 one-way traffic flows. Pontryagin's maximum principle was employed to develop an optimal control law that minimizes the total delay.

Further algorithms have been developed by Green (15), Longley (21), Gordon (13), and Ross et al. (37). Green considered 2 one-way traffic flows with random arrivals and compared the Gazis-Potts algorithm with a modified vehicle-actuated algorithm. Longley considered the problem of unblocking the secondary roads when the primary roads are oversaturated. For fixed cycle time, the split is determined by current queue length and clearance of as many secondary intersections as possible. Gordon considered the same problem as Longley and used Z-transform analysis for the development of a control algorithm. The stability condition was also derived. The work by Ross et al. considered not only the minimization of the delay of the critical intersection but the minimization of the downstream intersection delay as well. Their method is based on the control concept of Miller (25) and Weinberg et al. (48), wherein

a computer-control algorithm was developed for real-time control and simulation results were presented.

All those research efforts on single-intersection control seem to indicate that little is left for further investigation except the oversaturated traffic condition and the strategic control. We could study, for example, under what circumstances is a fixed-cycle control better than an actuated control and vice versa. The results could be helpful to traffic engineers in selecting different control strategies.

Signal Settings of an Arterial

Most researchers uphold the minimum delay criterion of signal settings at a single intersection but do not do so for an arterial. In this case, there is no satisfactory expression for delay—at least for the present—that links delay, cycle time, splits, and other necessary traffic parameters, so that the minimization process can be carried out in a manner similar to that for a single intersection. The difficulties arise from 2 sources: the increasing number of variables and the nonindependent arrivals due to the short distance between each pair of adjacent intersections.

Morgan and Little (26) have developed an algorithm that is based on the criterion of maximum bandwidth. Based on their definition, the bandwidth along a street where there is a sequence of signals all of which have the same cycle time is that portion of a cycle during which a car could start at one end of the street and, traveling at a preassigned speed, go to the other end without stopping for a red light. Basically, their algorithms do the following:

1. Synchronize the signals to produce bandwidths that are equal in each direction and as large as possible, given an arbitrary number of signals, a common signal cycle, the green and red times for each signal, and specified travel times between adjacent signals; and
2. Adjust the synchronization to increase one bandwidth to some specified, feasible value, and maintain the other as large as is then possible.

A mixed-integer linear program formulated by Little (20) calculates the common cycle time, speed between signals, and offsets to maximize the sum of the bandwidths for the 2 directions, given an arbitrary number of signals, the red-green split for each signal, upper and lower limits on signal period, upper and lower limits on speed between adjacent signals, and limit on change of speed.

The algorithms given above obviously assume that the platoon moves at a constant speed between each pair of adjacent intersections and that traffic does not turn either in or out. Besides these shortcomings, an algorithm that maximizes bandwidth may or may not yield minimum total delay. This is particularly true when the number of intersections along an arterial is large and irregularly spaced. Under such circumstances, the maximum equal bandwidth is likely to be zero for a given cycle time and split.

The algorithm developed by Yardeni (50) is similar to the maximum bandwidth design in many respects. Yardeni's concept is probably the same as that of Morgan and Little; that is, it is desirable to have nonstop travel along an arterial. However, the approach is different in that it is not trying to maximize bandwidth; instead it defines a generalized least squares fit criterion for the minimization. He observed that the ratio of the bandwidth to the cycle time depends only on the product of the cycle time and the design speed. The least squares process is to set the round-trip travel time between each pair of intersections at the design speed as close as possible to an integer multiple of the common cycle time. When the common cycle time is given, offsets are chosen to give the best feasible throughband. Although the performance of Yardeni's algorithm has not been thoroughly evaluated, his offset design is less effective than those of the others, for instance, the design by Morgan and Little (26) and the combination method by Hillier (16). This fact is observed by Wagner et al. (45).

Newell (31) also analyzed the 2-way arterial problem. In his analysis he made the following assumptions:

1. There is no turning traffic, and the same number of vehicles pass each intersection in each cycle; and

2. The platoon released at one intersection is still accelerating when it reaches the next and has not caught up with the tail of the preceding platoon, and the flow is high enough for the arrivals at each intersection to be spread over a period longer than the effective green time.

He showed that the delay is minimized for one direction by an offset that allows the last vehicle in the platoon arriving at any signal to cross the stop line just before the effective green time ends. For 2-way traffic, the offset that gives the least total delay between each pair of intersections is the one that gives the smaller minimum delay in either of the 2 directions. A second analysis by Newell (33) considered high flow of traffic in an arterial such that the cycle time and the effective green portion are the same for all intersections along the arterial. Furthermore, if we let S_1 and S_2 be the 2-direction saturation flow rates at each intersection, it is assumed that the flow is so high that, throughout the entire effective green period, the departures are at the saturation rates S_1 and S_2 for the 2 directions. Newell showed under the assumptions given above that, if there is no turning traffic and travel speeds are the same, the delay and the number of stops are both minimized by the offset on the link with the higher saturation flow being made equal to the travel time on that link for the case $S_1 \neq S_2$. For $S_1 = S_2$ there can be a range of offsets that minimizes both the delay and the number of stops, and we can therefore select the offset as either t (the travel time between the 2 intersections of the link) or C - t—they gave identical results.

The results given above are deceptively simple, for traffic rarely satisfies the assumptions. None of the algorithms given above has considered the left-turn or right-turn effects, optional lane effects and multiphase signalization. However, in a study of the signal control of a diamond interchange arterial by the authors (28), some of the algorithms have been extended to include these important factors. It is observed that no single algorithm gives uniformly better results than any of the others for different traffic flow levels and patterns.

Yagoda et al. (49) have developed for 2-way arterial signal settings a new technique that allows the cycle time to be variable and uses the excess capacity available at the various intersections along the arterial to obtain best progression. An advantage of this method is the relatively simple computation procedure. Again, the method does not give the minimum of a linear combination of the total delay and stops on the arterial.

Other control concepts and algorithms developed for networks can also be applied to arterials if arterials are considered as special networks. They are discussed in the next section.

Signal Settings of a Network

There are, in general, 4 types of networks: tree, ladder, tree of ladders, and general grid.

A single intersection and an arterial are special cases of a tree, the simplest type of network. The results of Newell (33) immediately apply to this case, provided the assumptions stated in the previous section still hold for each link.

Algorithms particularly developed for tree networks are those by Hillier (16) and by Inose et al. (19). In the work by Hillier, networks are considered as graphs, intersections as vertexes, and links as axes. A ladder or a tree of ladders can be successively simplified into a single link or tree by combining pairs of axes in parallel and in series. The calculation of delay used the assumption that the flow in each link is considered uniform, and 2 flow levels correspond to the effective green and red periods. The objective function is the minimization of total delay in each link. The total delay per cycle on each link is assumed to be a function of the offsets of the signals at the ends of this link, for a given cycle time and splits, and independent of the other offsets in the network. An integer N is also given, so that the offset can only be integral multiples of cycle C divided by N. The algorithm then calculates the sum of 2-direction delay by varying the offset iC/N, i = 0, 1, 2, . . ., (N - 1), and selects the one that gives minimum 2-direction delay. This is known as the delay-difference method. This method appears time-consuming for calculating the offset for a single link but has an advantage in that the computation increases linearly with the number of links in the

network, as compared with other methods in which computations generally increase nonlinearly.

Allsop (1) extended Hillier's work to cover general grid networks, i.e., networks that cannot be simplified to a tree by successive serial and parallel combinations. An algorithm that minimizes total delay and gives global minimum was developed later by Allsop (3), using a dynamic programming approach, and was written in FORTRAN. Unfortunately, the computer running time is too long for a medium-sized network (50 intersections), and the algorithm cannot be used for real-time computer control. One shortcoming of the combination method is the assumption of independence of the delay-difference function from the offset. The other is that it requires excessive computer core storage.

The algorithms developed by Inose et al. (19) applied only to tree networks. The cycle time of this network is taken as the maximum of the individual intersection ideal cycle time. The individual intersection ideal cycle time is one that minimizes the intersection delay time by considering the intersection as being isolated and having a uniform arrival rate. Once the common cycle is chosen, the splits are also obtained by minimizing the average delay with uniform arrival rates. Offsets are obtained by simplifications that neglect turning traffic and assume constant vehicle speed between intersections. The offset is chosen in each link as the travel time corresponding to the direction with higher flow. In a general network, this can be freely chosen provided that the links do not form any closed loops. They, therefore, choose a maximal tree in the graph. If the network has n vertexes, the maximal tree is the first n - 1 links that give the most delay saving, when we minimize the delay of each link. Let us identify each link in decreasing order, selecting them one by one and assigning the ideal offset to each. If a link happens to close a loop, it is discarded and the next one is considered, until all vertexes are covered. This method is simple and effective if there is little or no turning traffic. However, it is unlikely that this choice of maximal tree gives minimum total delay of the network.

Weinberg at al. (48) use the same principle as used for an isolated intersection to determine the signal settings of a network. Once again the decision whether to extend the green time or to switch depends on the estimated difference in delay for the competing approaches. The additional information needed in this case includes the estimation of the nearby signal settings. There are 2 drawbacks to their approach. First, the nearby intersection signal settings have to be estimated. It is unlikely that these estimates will be very accurate, because they are dependent variables of the current optimization. Second, it is assumed that the delay at one intersection is independent of delays at all other intersections, except those immediately nearby. Their algorithms have not been programmed yet. The effort involved in programming these algorithms is expected to be extensive.

SIGOP developed by Traffic Research Corporation (40) is probably the first systematic computer optimization program for a general network. For any given cycle length, the program first computes the phase split for each intersection. The length of each phase is set proportional to the total flow or critical flow in each phase or any desired combination thereof. Total flow is the sum over all lanes and all approaches, and critical flow is the maximum flow observed per lane through an intersection. To assist traffic engineers in selecting the best cycle length for this system, the program will optimize and evaluate up to 10 given cycle lengths in a single computer run. The optimization procedures include the determination of ideal offset differences for every link and the computation of optimal offsets for the entire network.

For a given link, the offset difference is the difference in time between the beginning of green at the downstream and upstream signals. Some ideal offset difference exists that minimizes delay or stops or a linear combination thereof. From the work by Newell (31), it seems desirable to stop the head of a platoon for a short period of time and allow the last vehicle of a platoon to clear. If, as Inose et al. assumed, there is no platoon dispersion, then the ideal offset difference is just the travel time, provided the green time is about the same for upstream and downstream signals. The exact platoon dispersion and its stochastic behavior affect the true ideal offset differences.

In general, there does not exist a set of offset differences for the entire network such that all links have their ideal offset differences. SIGOP is to find a set of optimal offset differences that minimize a weighted sum of squares of the differences between the ideal and optimal set for each link. Different weighting factors can be assigned to each link, according to their importance, which can be specified by the traffic engineer. More specifically, the delay at link ij per cycle is approximated by

$$f_{ij} = A_{ij}(R_{ij} + M_{ij} + d_i - d_j)^2 + S_{ij} \tag{11}$$

for i, j = 1, 2, . . ., n, and for an n-node intersection network,

where

i is the upstream intersection;
j is the downstream intersection;
d_i is the offset of intersection i;
R_{ij} is the ideal offset difference for the link;
A_{ij} is the link weighting factor;
S_{ij} is a constant;
f_{ij} is the delay for the link; and
M_{ij} is an integer chosen such that $-\frac{1}{2} \leq R_{ij} + M_{ij} + d_i - d_j < \frac{1}{2}$, all d_i and R_{ij} are expressed in fractions of a cycle.

SIGOP is to find d_i, i = 1, 2, . . ., n such that

$$F = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij}(R_{ij} + M_{ij} + d_i - d_j)^2 + S_{ij} \tag{12}$$

is minimized.

If all d's are held constant except d_i and d_j, M_{ij} can be shown to have, at most, 3 distinct values: minimum M^1, $M^1 + 1$, and $M^1 + 2$. Each M_{ij} is just a hyperplane, and they form parallel planes for each pair of intersections. Each group of hyperplanes is intersected by every other hyperplane. It is shown that the entire space is divided into 3^p by 5^q subregions, where p is the number of 1-way links, and q is the number of 2-way links. It seems a formidable task to find the global minimum by examining the local minimum of each subregion. However, in SIGOP, it can be seen that, of the local minimum in one subregion is found, all other neighboring subregion minimums can be found. The search for the global minimum starts at the local minimum of a randomly chosen subregion and transforms to the least of the adjacent local minimums. This process is continued until a subregion is found where the local minimum is less than that of any adjacent subregion minimum and is taken as a first estimate of the global minimum. This process is repeated with new randomly selected starting points and new estimates of the global minimum, comparisons are made with the old minimum, and the smaller minimum is maintained. This process terminates if any of the 10 estimates do not give a smaller minimum value than the old estimated global minimum.

Mathematically, we can see that SIGOP does not guarantee that the global minimum will be found; moreover, unless we start at a good initial point, a huge number of estimates may be necessary to terminate the computation. Furthermore, an inversion of n by n symmetric matrix is needed during the computation, where n is the number of intersections. This becomes a time-consuming task if n is large. Besides these shortcomings, the assumption by SIGOP that the ideal offset difference of each link is independent of cycle time, as it uses up to 10 different cycle lengths to compute optimal offsets without recomputing R_{ij} each time, is likely not valid. It is also observed that f_{ij} is only a very crude estimate of the delay except when the computed offset differ-

ence is very close to the ideal offset difference. Nevertheless, SIGOP has been shown to be an improvement over existing signal settings and compared favorably with other methods (Wagner et al., 46).

Unlike the algorithms mentioned above, the method by Chang (1) regards both the green phase splits and the offsets as variables. He assumes that traffic arrives in platoons but is uniform within the platoon and that traffic enters each link from only one approach at the upstream signal. He then obtains equations determining the queue length for any signal settings. The objective function is

$$F = (1/T) \sum_{i=1}^{n} \int_{0}^{T} q_i(t)dt \tag{13}$$

where T is the cycle time, and $q_i(t)$ is the steady-state queue length at intersection i at time t. A gradient technique is developed to minimize F over all possible offsets and splits. However, it is suspected that the gradient search will provide only a local minimum of F.

A more sophisticated model and a more effective optimization technique are employed in TRANSYT by Robertson (36). This model has been described earlier in this paper. TRANSYT makes the following assumptions:

1. All major junctions of the network have signals (or are controlled by a priority rule);
2. All signals in the network have a common cycle time or a cycle time half this value;
3. Traffic enters the network at a constant specified rate on each approach; and
4. The platoon of traffic turning left or right at each signal remains constant throughout the cycle.

The objective function is a linear combination of delay and stops. The common cycle time at the signals is divided into 50 equal units of time. All TRANSYT calculations are made on the basis of the average values of flow rates and vehicle queues that are expected to occur during each of these units of time. In practice, the random behavior of individual vehicles will vary the average flow pattern; a smoothing process is therefore employed to give average platoon shape. They observe that a cycle time that is long enough to clear the queue within one green period for uniform arrival rate may not be sufficient for random arrival for every cycle. The extra delay per car caused by the random behavior of the platoon is found to be

$$D = X^2/[4d(1 - X)] \tag{14}$$

where X is the degree of saturation, and d is the average arrival rate.

It is interesting to note that Eq. 14 is exactly one-half of the second term of Webster's delay equation (Eq. 7). Webster claims that his second term is due to the randomness of the arrival. In Webster's derivation, an intersection is considered as being isolated and arrivals as being independent, identically distributed random sequences. This is no longer true for platoons: It results in the reduction to half of Webster's observation. It is also interesting to point out that Webster's equation for determining cycle length is generally too long for arterials or networks for the same reason that delay is overestimated, and Inose et al. (19), or all others who assume uniform arrival rate to determine optimal cycle length, generally specify too short a cycle time in that they have underestimated the delay.

A computational procedure called hill-climb is used by TRANSYT to find the minimum of the objective function. The number of computations involved increases in the order of the square of the number of intersections. This optimization procedure seems more

encouraging than that of SIGOP, which involves the inversion of a symmetric matrix of order of the number of intersections and a space divided into a tremendous number of subregions. However, the same difficulty occurs in TRANSYT as in SIGOP: We may end up with a local minimum instead of the global minimum of the objective function.

TRANSYT has been field-tested in London and Glasgow and has shown better results than the combination method. SIGOP, the maximum bandwidth, and the combination method have been compared by simulation and field tests by Wagner et al. (46). In their study, cycle time is allowed to vary. There is no indication that one method is uniformly better than the other. There is also a plan to test TRANSYT and compare it with SIGOP in some American cities of which San Jose, California, is one of the candidates. SIGOP has also been field-tested in Kansas City, Missouri, where improvements were reported on both travel time and total stops over existing systems. Five other cities are currently ready to implement the program. SIGOP is also to be tested in Glasgow, Scotland, where a comparison between TRANSYT and SIGOP will be made.

Inasmuch as the most complete and systematic network control algorithms are the combination method, SIGOP, and TRANSYT, we shall give a summary of them in Table 1 for comparative purposes.

We also need to know that the 3 algorithms are all for fixed-time control and that computing is generally too long for real-time computer control.

CURRENT DEVELOPMENTS IN TRAFFIC NETWORK CONTROL

A series of urban network control research programs are currently in progress for the Federal Highway Administration. They are generally known as the First Generation, the Second Generation, and the Third Generation Control Software Programs. The First Generation Control Software Program was conducted at Sperry Rand (38). The major tasks were to establish traffic parameters to be used as the basis for on-line selection and evaluation of control strategies and to develop concepts for a traffic signal control system that can readily implement experimental control strategies. The traffic parameters selected consist of occupancy, volume, queue, stops, delay, speed, and travel time—the first four are used in the algorithms that determine the traffic signal settings, and the last four measure system effectiveness. These parameters are obtained either from direct detector measurements or from a set of equations that assume uniform platoon behavior with a set of deceleration characteristics. The control strategy is basically table-look-up type, with stored historical data. Table selection could be achieved by a calculation of the sum of squares of the differences between certain measured traffic parameters (e.g., volume, speed, and occupancy are suggested by the Department of Transit and Traffic, City of Baltimore, 4) and those that would result in the signal settings in the table and a selection of the one that gives the minimum. Signal timing can be switched every 15 min—if the new traffic differs significantly from any of the stored histories for a set of controllers, SIGOP is then employed off-line to generate new timing plans based on the new traffic data. Critical intersection control is also provided, where a critical intersection is one that frequently oversaturates. When critical intersection control is necessary, the split at a critical inter-

Table 1. Comparison of major network-control algorithms.

Algorithm	Model	Prediction	Objective	Optimization Technique	Network
Combination	Uniform	None	Minimizing delay	Delay-difference (dynamic prog.)[a]	Tree, ladder, tree of ladders (any)[a]
SIGOP	Uniform	None	Minimizing least squares deviation from ideal offsets	Gradient	Any
TRANSYT	Stochastic	Dispersion of platoon	Minimizing linear combination of delays and stops	Hill climb	Any

[a]Parentheses give Allsop's extensions to Hillier's work.

section is set proportional to the smoothed queue length ratio. Bus priority control is also implemented at certain critical intersections. The decision whether to grant additional green time for a particular phase depends on the queue ratio as well as on the detection of a bus on the approach to the intersection.

From the discussion given above, we understand that the First Generation Control Software Program employs a fairly simple model to describe traffic dynamics, and the optimization of signal timings depends on SIGOP. Therefore, it is not a true minimization of delay.

The Second Generation Control Software Program is currently being developed by TRW. The primary objective is to provide on-line generation of optimal signal settings. Once again, SIGOP is used—with minor modifications—for optimization of offsets. The Second Generation Control Software Program differs from the First Generation Control Software Program in that it requires network decomposition so that the optimization time of each subnetwork by SIGOP is short enough compared to the 15-min signal timing setting. Signal settings are modified every 15 min according to the measured and predicted traffic data for the next 15 min. A predictor that can project traffic conditions for the next 15-min period is necessary to achieve this task. The criteria for subnetwork decomposition are those proposed by Walinchus (43). Once again, these decomposition criteria rely on the least squares principle, whose relations to the minimum delay are not clear.

The near real-time (15 min) computer control gives better traffic operation than that given by all existing computer control systems, it is claimed by TRW in its SAFER (Systematic Aid to Flow on Existing Roadways) system. It is also demonstrated by the authors (29) that near real-time operation yields better results than multidial systems for a closed-loop network.

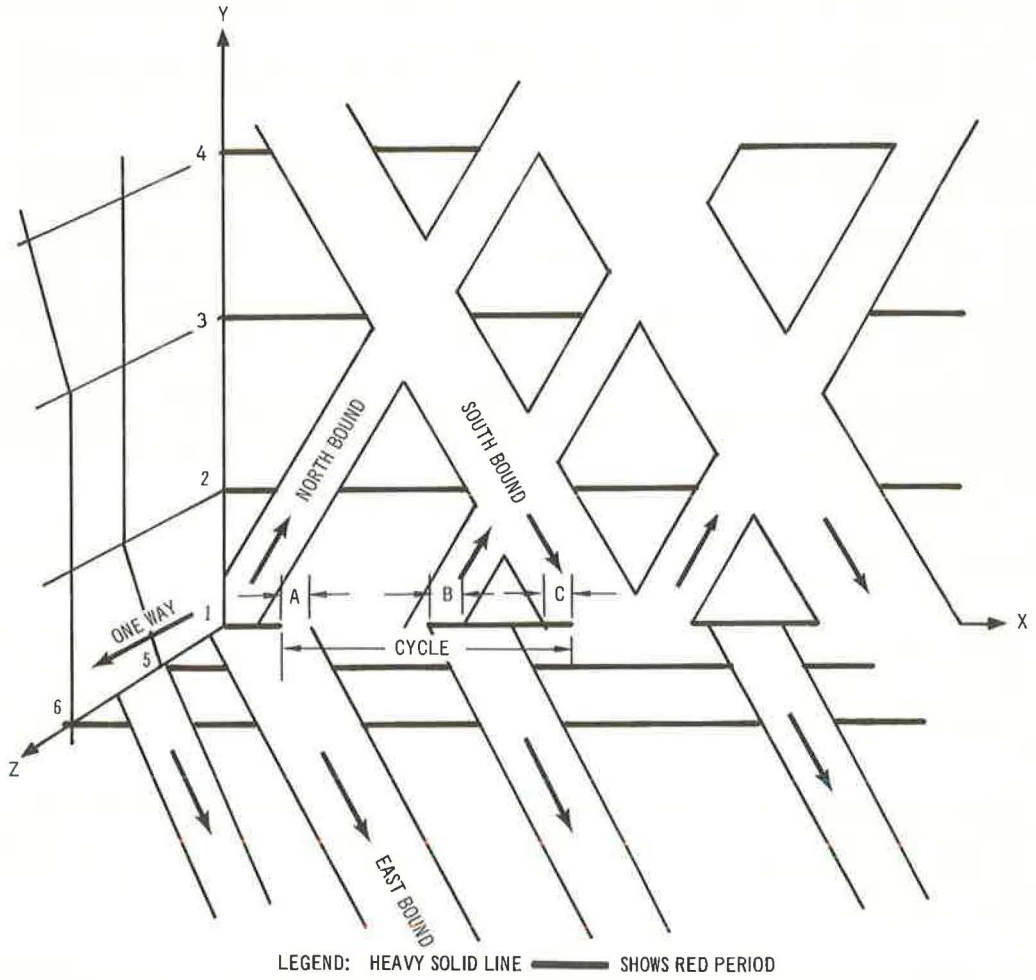
Under general conditions, using a common cycle time for the entire network may not yield better results than employing uncommon cycle times. This leads to the research of the Third Generation Control Software Program (U. S. Department of Transportation, 41), which can be characterized by the dynamic optimization of traffic signal settings without the up-to-now imposed requirement of a background cycle. This results in the development of a variable-cycle signal-timing program.

There are perhaps only 2 dynamically responsive control algorithms that have appeared in the literature. The first, proposed by Miller (25) and developed by Weinberg et al. (48), has been discussed earlier; it minimizes total delay n seconds into the future, where n is small. This procedure is questioned because it records immediate gain without analysis of the consequence of the action. In an extensive simulation of a fairly elaborate intersection at a variety of flow levels by Van Zijverden et al. (42), a minimized short-term delay strategy provided no better results than a good traffic-actuated strategy. For these reasons, algorithms by Miller and Weinberg at al. are considered at this time to be of dubious value.

The second dynamically responsive control algorithm, PLIDENT by Holroyd and Hillier (18), was field-tested in Glasgow, Scotland. This scheme has neither fixed-cycle nor offset concepts, but adjusts signal settings to suit various platoons. Platoons on main roads are identified, and their arrival times at downstream signals are predicted. Despite the elaborate PLIDENT algorithm, it produces very poor results in terms of travel time, compared to those of TRANSYT and the combination methods. One clear fault is the way in which cross-street traffic was not considered in the calculation of the signal settings. Furthermore, PLIDENT does not minimize delay, and the other 2 methods do.

The Third Generation Control Software Program, using the same concept of no fixed-cycle time and offsets, approaches the optimal signal setting in a mathematically delicate and rigorous way. This can be illustrated by the 3-dimensional time-space diagram shown in Figure 1. Let us consider intersection 1, where the X-axis represents the history of the north-south signal and the Z-axis represents the history of the east-west signal. The objective is to minimize a linear combination of delay and stops over all intersections and over the entire time period of interest. If the platoon is considered uniform, i.e., no compression and dispersion of the platoon, then delay and

Figure 1. Three-dimensional time-space diagram.



stops can be expressed during each cycle as functions of the platoon conflicts A, B, and C. More specifically,

$$d \simeq A \cdot q_E \cdot [T_1 + (A/2)] + B \cdot q_N \cdot [T_2 + (B/2)] + C \cdot q_S \cdot (C/2) \quad (15)$$

$$s = A \cdot q_E + B \cdot q_N + C \cdot q_S \quad (16)$$

where A, B, and C are respectively eastbound platoon, northbound platoon, and southbound platoon to be stopped; q_i is the corresponding platoon density, i = E, N, S; T_1 is the east-west effective red period; T_2 is the north-south effective red period; d is the total delay suffered by the 3 platoons at intersection 1; and s is the number of stops of the 3 platoons at intersection 1.

The objective is to minimize a linear combination of d and s over all intersections and over the entire time period of interest. To do this requires that an accurate mathematical model first be developed. The sophistication of the model depends on how

platoon behavior is defined. Inasmuch as the direction is toward increasing sophistication, the platoon can be uniform and travel can be at a constant speed, it can be uniform with increased complexity of compression and dispersion characteristics, or it can be of a stochastic nature with random fluctuations. This mathematical model, then, serves as a basis for predicting traffic patterns for the next T min, where T is fairly large. Accurate prediction techniques will also be developed to obtain predictions of necessary traffic data for optimization. The optimization program employs the predicted data and calculates optimal signal settings (or suboptimal, if the network has to be decomposed). The optimal signal settings are followed for a short period of time, after which the predictor updates its prediction and the optimization processor reoptimizes signal settings. This procedure is repeated for the entire time period of interest.

This seems plausible. The key point for success is the speed of the optimization program. If it is too slow, say, 5 min, then the signal settings must follow the calculated optimums for at least 5 min before the optimization program can reoptimize by using updated traffic data. The error rate is likely to be high in this case. Therefore, the network may need to be decomposed into subnetworks to achieve higher optimization speed. SIGOP, used by the First and Second Generation Control Software Programs, is definitely too slow; furthermore, it fails to give minimum delays or stops. The Third Generation Control Software Program is a new control concept. If successful, the problem of traffic control of networks can be greatly benefited.

Two major problems that are of vital importance in terms of real-time control and have not received much attention in the research literature are the transition and reconciliation problems. The former has its role in first and second generation control in that, when a new signal plan is determined, there should be a smooth transition from the old signal plan to the new one so that traffic disturbance is kept minimal. The latter has its role in second and third generation control in that, when a network is decomposed into subnetworks, some mechanism is needed to reconcile signal settings between subnetworks so that traffic disruption is minimal in the boundaries and, at the same time, to achieve some degree of optimality (if not true optimal) of each subnetwork. Walinchus (43) has addressed the reconciliation problem and suggested some approaches to solutions. However, we feel that adequate optimization criteria should be defined, so that we can systematically solve the transition and reconciliation problem. We mention a possible way to achieve transition in the following.

Within each subnetwork, we can add a constant to the offset of each intersection; and the relative offset differences between each pair of intersections are unaffected. Only the offset differences between the intersection in the subnetwork and the master clock are affected. By adding a constant to the newly calculated offsets, we may achieve minimal disturbance within the subnetwork. That is, we find x for an n-intersection subnetwork such that

$$y = \sum_{i=1}^{n} (\theta_i - \phi_i - x + k_i C)^2 \qquad (17)$$

is minimized, where

C = new cycle time for next signal timing period;
θ_i = old offset for intersection i with respect to C, $0 \leq \theta_i < C$;
ϕ_i = new offset for intersection i, $0 \leq \phi_i < C$; and
k_i = an integer such that $-(C/2) < \theta_i - \phi_i - x + k_i C \leq (C/2)$.

Let x_o be the minimizing value of x. The addition of x_o to θ_i for each i minimizes the least squares deviation between the old and the new offsets. If there is more than one subnetwork, we must reduce their differences to zero gradually so that all subnetworks are synchronized. This can be achieved by increasing or decreasing the cycle time of some subnetworks by 1 or 2 sec each cycle until synchronization. We credit this concept to the City of Baltimore Department of Transit and Traffic (4).

Dynamic programming is also an encouraging technique for optimizing the objective functions. The works by Okutani (35) and Gartner (10, 11) belong to this category. Okutani considers a general network with common cycle time and splits given. The objective function of a link is assumed to depend only on the flow on this link, and the offset difference is independent of the offset differences of other links. The mathematical principle is that the optimization problem is considered a multistage decision process, and the objective function is minimized in passing through each stage.

The work of Gartner, in many respects, is similar to that of Okutani. However, Gartner uses graph theory to simplify the network. By doing so, he greatly reduces the number of intersections in the simplified network. This makes the optimization process more feasible if a real-time computer control operation is desired because the computer running time increases only linearly with respect to the number of nodes of the simplified network. Although both authors have used simplified traffic models like the fluid model and their results have not been compared with other control algorithms, we believe that the dynamic programming technique may yield more fruitful results in traffic network control in the near future.

SUMMARY

We have discussed major traffic control concepts and algorithms, reported in recent published literature, in terms of the mathematical model, the prediction technique, and the optimization of the objective function.

For an isolated intersection, we have described several arrival models. In each case, an expression for delay is given that minimizes the expected average delay. Different arrival models result in different expressions for the average delay. Optimal control schemes are summarized for both vehicle-actuated controllers and fixed-cycle controllers. For vehicle-actuated controllers, the optimal scheme is to switch signals as soon as the queue is cleared. For fixed-cycle controllers, the optimal scheme is, in general, to keep the cycle time as small as possible while the demand is being observed.

For the control of an arterial, the mathematical models used usually involve the description of platoon behaviors. Uniform and stochastic models, with or without compression or dispersion, are both discussed in the literature. The objective function is usually to maintain progression, to maximize the bandwidth, or to obtain best least squares fit.

For the control of a network, a wide variety of models are proposed. Many algorithms require the prediction of traffic parameters and patterns. The objective function is usually the minimization of delay or a linear combination of delays and stops. In either case, the optimization procedure is extremely complicated, and even sophisticated computational methods do not give the global minimum of the objective function. If real-time or near real-time control is desired, breakdown of the network into subnetworks is proposed to decrease computer running time necessary for calculating optimal signal settings.

Current trends in the research of network control are also discussed.

REFERENCES

1. Allsop, R. E. Selection of Offsets to Minimize Delay to Traffic in a Network Controlled by Fixed-Time Signals. Transportation Science, Vol. 1, 1967, pp. 1-13.
2. Allsop, R. E. An Analysis of Delays to Vehicle Platoons at Traffic Signals. Proc. Fourth Internat. Symposium on Theory of Traffic Flow, Karlsruhe, Germany, 1968.
3. Allsop, R. E. Optimization Techniques for Reducing Delay to Traffic in Signalized Road Networks. University College, London, PhD thesis, 1970.
4. Preliminary Technical Specifications for the City of Baltimore, Traffic Command and Control System. Department of Transit and Traffic, City of Baltimore, 1971.
5. Bavarez, E., and Newell, G. F. Traffic Signal Synchronization for High Flows on a One-Way Street. Transportation Science, Vol. 1, 1967, pp. 55-73.
6. Beckmann, M. C., McGuire, B., and Winsten, C. B. Studies in Economics of Transportation, Yale Univ., 1956.

7. Chang, A. Synchronization of Traffic Signals in Grid Networks. IBM Systems Jour., Vol. 6, 1967, pp. 436-441.
8. Clayton, A. J. H. Road Traffic Calculations. Jour. of Institution of Civil Engineers, Vol. 16, No. 7, 1941, pp. 247-284.
9. Gafarian, A. V., Hayes, E., and Mosher, W. W., Jr. The Development and Validation of a Digital Simulation Model for Design of Freeway Diamond Interchanges. Highway Research Record 208, 1967, pp. 37-78.
10. Gartner, N. Optimal Synchronization of Traffic Signal Networks by Dynamic Programming. Fifth Internat. Symposium on Theory of Traffic Flow and Transportation, Univ. of California, Berkeley, 1971.
11. Gartner, N. Algorithms for Dynamic Road Traffic Control. Fifth IFAC World Congress, Paris, June 1972.
12. Gazis, D. C. Optimal Control of Oversaturated Intersection. Proc. Second Internat. Symposium on Theory of Traffic Flow, London, 1964.
13. Gordon, R. L. A Technique for Control of Traffic at Critical Intersections. Transportation Science, Vol. 3, No. 4, 1969, pp. 279-288.
14. Grafton, R. L., and Newell, G. F. Optimal Policies for the Control of an Undersaturated Intersection. Proc. Third Internat. Symposium on Theory of Traffic Flow (Edie, L., ed.), American Elsevier Pub. Co., New York, 1967
15. Green, D. H. Optimal Control of a System of Oversaturated Intersections. Operations Research, Vol. 12, 1967, pp. 815-831.
16. Hillier, J. A. Appendix to Glasgow's Experiment in Area Traffic Control. Traffic Engineering and Control, Vol. 7, No. 9, 1966, pp. 569-571.
17. Hillier, J. A., and Rothery, R. The Synchronization of Traffic Signals for Minimum Delay. Transportation Science, Vol. 1, 1967, pp. 91-94.
18. Holroyd, J., and Hillier, J. A. The Glasgow Experiment: PLIDENT and After. Joint Transportation Eng. Conf., Chicago.
19. Inose, H., Fujisaki, H., and Hamada, T. A Road Traffic Control Theory Based on a Macroscopic Traffic Model. J.I.E.E., Vol. 87, No. 8, 1967, pp. 55-67.
20. Little, J. D. C. The Synchronization of Traffic Signals by Mixed-Integer Linear Programming. Operations Research, Vol. 14, 1966, pp. 568-594.
21. Longley, D. A Control Strategy for a Congested Computer-Controlled Traffic Network. Transportation Research, Vol. 2, 1968, pp. 391-408.
22. Martin-Löf. Computation of an Optimal Control for a Signalized Traffic Intersection. Transportation Science, Vol. 1, Feb. 1967, pp. 1-5.
23. McShane, W. R., Yagoda, H. N., Pignataro, L. J., and Crowley, K. W. Control Considerations and Smooth Flow in Vehicular Traffic Nets. Highway Research Record 344, 1970, pp. 8-22.
24. Miller, A. J. A Queueing Model for Road Traffic Flow. Jour. of Royal Statistical Society, Series B, Vol. 23, 1961, pp. 64-75.
25. Miller, A. J. A Computer Control System for Traffic Networks. Proc. Second Internat. Symposium on Theory of Traffic Flow, London, 1963.
26. Morgan, J. T., and Little, J. D. C. Synchronizing Traffic Signals for Maximum Bandwidth. Operations Research, Vol. 12, 1964, pp. 896-912.
27. Munjal, P. K., and Fitzgerald, J. W. A Network Simulation Model of Traffic Evaluation of Computer Control Concepts. SDC Technical Memorandum, TM-4601/004/00, 1971.
28. Munjal, P. K., and Hsu, Y. S. Evaluation of Real-Time Control Algorithms for Diamond Interchange. SDC Technical Memorandum, TM-4601/003/00, 1971.
29. Munjal, P. K., and Hsu, Y. S. Some Simulation Results of Different Vehicular Traffic Control Strategies. IEEE/ORSA Joint National Conf. on Major Systems, Anaheim, Calif., IEEE Cat. No. 71C46-SMC, 1971, pp. 225-230.
30. Newell, G. F. Statistical Analysis of the Flow of Highway Traffic Through a Signalized Intersection. Quarterly Applied Mathematics, Vol. 13, No. 4, 1956, pp. 353-369.
31. Newell, G. F. Synchronization of Traffic Lights for High Flow. Quarterly Applied Mathematics, Vol. 21, No. 4, 1964, pp. 315-325.
32. Newell, G. F. Traffic Signal Synchronization for High Flows on a Two-Way Street. Proc. Fourth Internat. Symposium on Theory of Traffic Flow, Karlsruhe, Germany, 1968.

33. Newell, G. F. Properties of Vehicle-Actuated Signals: I. One-Way Streets. Transportation Science, Vol. 3, 1969, pp. 30-52.
34. Newell, G. F., and Osuna, E. E. Properties of Vehicle-Actuated Signals: II. Two-Way Streets. Transportation Science, Vol. 3, 1969, pp. 99-125.
35. Okutani, I. Synchronization of Traffic Signals in a Network for Loss Minimizing Offsets. Fifth Internat. Symposium on Theory of Traffic Flow and Transportation, Univ. of California, Berkeley, 1971.
36. Robertson, D. I. TRANSYT: A Traffic Network Study Tool. Road Research Laboratory, Crowthorne, Berkshire, England, Rept. LR 253, 1969.
37. Ross, D. W., Sandys, R. C., and Schlaefli, J. L. A Computer Control Scheme for Critical-Intersection Control in an Urban Network. Transportation Science, Vol. 5, No. 2, 1971, pp. 141-160.
38. Advanced Control Technology in Urban Traffic Control Systems. Sperry Systems Management Division, Sperry Rand Corp., Great Neck, New York, 5 vols., 1969.
39. Tanner, J. C. Delays on a Two-Lane Road. Jour. of Royal Statistical Society, Series B, Vol. 3, 1961, pp. 38-63.
40. SIGOP: Traffic Signal Optimization Program, A Computer Program to Calculate Optimum Coordination in a Grid Network of Synchronized Traffic Signals. Traffic Research Corp., New York, 1966.
41. Variable Cycle Signal Timing Program. Federal Highway Administration, U. S. Dept. of Transportation, RFP-173, 1971.
42. Van Zijverden, J. D., and Kwakernaak, H. A New Approach to Traffic-Actuated Computer Control of Intersections. Proc. Fourth Internat. Symposium on Theory of Traffic Flow, Karlsruhe, Germany.
43. Walinchus, R. J. Real-Time Network Decomposition and Subnetwork Interfacing. Highway Research Record 366, 1971, pp. 20-28.
44. Wardrop, J. G. Some Theoretical Aspects of Road Traffic Research. Proc. Institute of Civil Engineering, Vol. 1, No. 2, 1952, pp. 325-362.
45. Wagner, F. A., Gerlough, D. L., and Barnes, F. C. Improved Criteria for Traffic Signal Systems on Urban Arterials. NCHRP Rept. 73, 1969.
46. Wagner, F. A., Barnes, F. C., and Gerlough, D. R. Improved Criteria for Traffic Signal Systems in Urban Networks. NCHRP Rept. 124, 1969.
47. Webster, F. V. Traffic Signal Settings. Road Research Laboratory, Tech. Paper 39, London, 1958.
48. Weinberg, M. I., Goldstein, H., McDade, T. J., and Whalen, R. H. Digital Computer-Controlled Traffic Signal System for a Small City. NCHRP Rept. 29, 1966.
49. Yagoda, H. N., Stevens, E. E., and Zabala, E. A New Technique for the Optimal Timing of Two-Way Arterials. Joint Transportation Eng. Conf., Chicago, Oct. 11-14, 1970.
50. Yardeni, L. A. Algorithms for Traffic Signal Control. IBM Systems Jour., Vol. 4, 1970, pp. 148-161.

SPONSORSHIP OF THIS RECORD

GROUP 3—OPERATION AND MAINTENANCE OF TRANSPORTATION FACILITIES
Harold L. Michael, Purdue University, chairman

Committee on Traffic Flow Theory and Characteristics
Donald G. Capelle, Alan M. Voorhees and Associates, Inc., chairman
Patrick J. Athol, John L. Barker, Martin J. Beckmann, Martin J. Bouman, Kenneth A. Brewer, Donald E. Cleveland, Kenneth W. Crowley, Robert F. Dawson, Lucien Duckstein, Leslie C. Edie, H. M. Edwards, A. V. Gafarian, Denos C. Gazis, Daniel L. Gerlough, John J. Haynes, Edmund A. Hodgkins, James H. Kell, John B. Kreer, Leonard Newman, O. J. Reichelderfer, Richard Rothery, August J. Saccoccio, A. D. St. John, Willian C. Taylor, Joseph Treiterer, William P. Walker, Sidney Weiner, W. W. Wolman

K. B. Johns, Highway Research Board staff